

GenCore version 4.5  
(c) 1993 - 2000 Comm

Search time 25.08 Seconds

378.045 Million cell updates/sec

...HGQIDSMEQLVITYQPERKD 128

5

ers: 522463

meseqp/AI1967.DAT
meseqp/AI1987.DAT
meseqp/AI1989.DAT
meseqp/AI1990.DAT
meseqp/AI1991.DAT
meseqp/AI1992.DAT
meseqp/AI1993.DAT
meseqp/AI1994.DAT
meseqp/AI1995.DAT
meseqp/AI1996.DAT
meseqp/AI1997.DAT
meseqp/AI1998.DAT
meseqp/AI1999.DAT
meseqp/AI2000.DAT
meseqp/AI2001.DAT

ected by chance to have a  
of the result being printed  
score distribution.

Description
Amino acid sequence
Mature T-cell protein
Human TCR-1 polypeptide
Human polypeptide
Human polypeptide
Human TCR-1 polypeptide
Pseudomonas stutzeri
Potato starch bran
Potato class B starch
Protein encoded by
Protein encoded by

Accession	Gene	Protein	Length	Score	E-value	Search time	Updates/sec
1.5	Compugen Ltd.		12	10.1	16	25.08	2.0
1.5		IBDV GLS structural	12	10.1	16		
1.5		IBDV strain GLS 1a	12	10.1	16		
1.5		Infectious bursal	12	10.1	16		
1.5		Polyepitide encoded	12	10.1	16		
1.5		CDK2-green fluores	12	10.1	16		
1.5		Egline C-terminal	12	10.1	16		
1.5		Infectious bursal	12	10.1	16		
1.5		Ambo acid sequenc	12	10.1	16		
1.5		IBDV VP2-VP4-VP3	12	10.1	16		
1.5		Human protein clon	12	10.1	16		
1.5		C glutamicam prote	12	10.1	16		
1.5		Pineapple polyphen	12	10.1	16		
1.5		Human AFP protein	12	10.1	16		
1.5		Human protein sequ	12	10.1	16		
1.5		Pineapple polyphen	12	10.1	16		
1.5		Human protein sequ	12	10.1	16		
1.5		IBDV structural pr	12	10.1	16		
1.5		IBDV structural pr	12	10.1	16		
1.5		Consensus amino ac	12	10.1	16		
1.5		Human protein sequ	12	10.1	16		
1.5		Novel protein kina	12	10.1	16		
1.5		Protein having aut	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		
1.5		IBDV gene product.	12	10.1	16		
1.5		Sequence of wild t	12	10.1	16		
1.5		Segment A of IBDV	12	10.1	16		
1.5		IBDV polyprotein V	12	10.1	16		
1.5		IBDV VP2-VP3-VP4	12	10.1	16		
1.5		Amino acid sequenc	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		
1.5		Arabidopsis thaila	12	10.1	16		

ALIGNMENTS

RESULT	1
AAAB18763	
ID	AAAB18763 standard; Protein: 128 AA.
XX	
AC	AAAB18763;
XX	
DT	22-JAN-2001 (first entry)
DE	
XX	Amino acid sequence of the human Tc1-1b protein.
KW	Tc1-1; Tc1-1b; T cell malignancy; chromosome 14 abnormality; lymphoma
KW	T-cell leukemia; immunodeficiency syndrome; ataxia-telangiectasia.
XX	
OS	Homo sapiens.
XX	
PN	W0200055169-A1.
XX	
PD	21-SEP-2000.
XX	
PF	15-MAR-2000; 2000MO-US06612.
XX	
PR	15-MAR-1999; 99US-0124714.
XX	
FA	(UYUE-) UNIV JEFFERSON THOMAS.
XX	
PI	Croce CM, Pekarsky Y;
XX	
DR	WPI; 2000-611514/58.
XX	
XX	DR P-PSDB; AAF5822.
PT	Novel nucleic acid of Tc1-1 gene family, Tc1-1b, expressed in low
PT	levels in normal bone marrow and peripheral lymphocytes, but activated
PT	in T-cell leukemia and lymphoma, used to identify chromosome 14
PT	abnormalities -

XX Claim 7, Page 64; 70pp; English.  
 PS  
 CC The present sequence represents a human Tc1-1b protein. The Tc1-1b gene  
 CC is implicated in the development of T cell malignancies. Fragments of  
 CC Tc1-1b cDNA sequences are used for detecting a target sequence  
 CC indicating a chromosome 14 abnormality, such as a (14:14)(q11;q32)  
 CC translocation or a (14)(q11;q32) inversion. Tc1-1b antisense sequences  
 CC and antibodies are useful for treating a disease state such as T-cell  
 CC leukemia or lymphoma associated with a chromosome 14 abnormality. The  
 CC Tc1-1b gene and its gene product are useful for treating disease states  
 CC associated with the Tc1-1b locus on chromosome protein including  
 CC T-lymphocytic leukaemias, acute and chronic leukaemias associated  
 CC with the immunodeficiency syndrome ataxia-telangiectasia (AT).  
 CC  
 XX Sequence 128 AA;

Query Match 100.0%; Score 685; DB 21; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 1e-72;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASEASVRLGVPGRMLIQPGIYEDSGRTWTVVVFNPSPREMARASQSGSRPEPT 60  
 Db 1 maseasvrlgvpgrmlwqpglyedegrtwvvtvvrnpssrrekarasqsgsyepsit 60  
 QY 61 VHLWQKAVHTRELSSGQMPFSQLPAYWQLYGRKRYRADSFWEIADHGOIDSMEQLVL 120  
 Db 61 vhlwqkavhtrelssgqmpfsqlpaywqlyparkkyaadsrweiadhqidsmeqlvl 120  
 QY 121 TYPERKD 128  
 Db 121 typerkd 128

RESULT 2  
 AAR94976  
 ID AAR94976 standard; Protein; 108 AA.  
 AC AAR94976;

DT 23-AUG-1996 (first entry)  
 DE Mature T-cell proliferative 1 protein.

XX Tc1-1; chromosome-14; leukaemia; lymphoma; therapy; diagnosis;  
 KW T-cell proliferative 1 protein.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FT Misc-difference 108 /note="unidentified amino acid"  
 FT

XX WO9613514-A1.

XX 09-MAY-1996.

XX 23-OCT-1995; 95WO-US13663.

XX 27-OCT-1994; 94US-0330272.

XX (RAGG-) RAGGIO-ITALGENE SPA.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Croce CM, Russo G;

XX WPI; 1996-239444/24.

XX New Tc1-1 protein and gene associated with chromosome 14  
 PT abnormalities - useful to develop prods. for detection, treatment  
 CC and prevention of diseases such as T-cell leukaemia(s) and  
 PT lymphoma(s)

XX Disclosure; Page 69; 105pp; English.  
 PS  
 CC The mature T-cell proliferative 1 protein (AAR94976) is involved  
 CC in a translocation of chromosome 14 and X chromosome  
 CC t(14:X)(q11;q28). It shows 408 amino acid sequence homology to  
 CC human Tc1-1 protein (AAR94974), which is associated with  
 CC chromosome 14 abnormalities leading to leukaemia and lymphoma.  
 CC  
 XX Sequence 108 AA;

Query Match 25.8%; Score 176.5; DB 17; Length 108;  
 Best Local Similarity 33.6%; Pred. No. 4.2e-13;  
 Matches 43; Conservative 17; Mismatches 43; Indels 25; Gaps 4;

QY 1 MASEASVRLGVPGRMLIQPGIYEDSGRTWTVVVFNPSPREMARASQSGSRPEPT 57  
 Db 1 maseasvrlgvpgrmlwqpglyedegrtwvvtvvrnpssrrekarasqsgsyepsit 57  
 QY 58 SITVHLWQKAVHTRELSSGQMPFSQLPAYWQLYGRKRYRADSFWEIADHGOIDSMEQ 117  
 Db 58 s---hllt-----sqldmwllypeerymdnstrlwlqldhlmvrygqe 98  
 QY 118 LVITYQPE 125  
 Db 99 lllkllpd 106

RESULT 3  
 AAR94974  
 ID AAR94974 standard; Protein; 113 AA.  
 AC AAR94974;

DT 23-AUG-1996 (first entry)

DE Human Tc1-1 polypeptide.

XX Tc1-1; chromosome-14; leukaemia; lymphoma; therapy; diagnosis.  
 KW  
 XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 5 /label= Phosphorylation site  
 FT /note="casein kinase II phosphorylation site"  
 FT

XX WO9613514-A1.

XX 09-MAY-1996.

XX 23-OCT-1995; 95WO-US13663.

XX 27-OCT-1994; 94US-0330272.

XX (RAGG-) RAGGIO-ITALGENE SPA.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Croce CM, Russo G;

XX WPI; 1996-239444/24.

XX N-PSDB; AAT18876.

XX New Tc1-1 protein and gene associated with chromosome 14  
 PT abnormalities - useful to develop prods. for detection, treatment  
 CC and prevention of diseases such as T-cell leukaemia(s) and  
 PT lymphoma(s)

PS Claim 6; Page 66; 105pp; English.

XX The amino acid sequence of human Tc1-1 protein (AAR94974) was  
 CC deduced from a cDNA clone (AAT18876) obtd. from an ALL cell line cDNA



DR N-PSDB; AAI60765.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX Example 2: SEQ ID NO 6540; 10078bp; English.  
 PS  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAI38642-AAI42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 XX Sequence 129 AA:  
 SQ  
 Query Match 21.8%; Score 149; DB 22; Length 129;  
 Best Local Similarity 29.2%; Pred. No. 9, 1e-10;  
 Matches 33; Conservative 21; Mismatches 45; Indels 14; Gaps 1;  
 QY 13 PGRMTIORPGIYDEDEGRVTWTVVVRNPSSRRMARASQGSRYESTIVHMQMAYHTE 72  
 Db 30 pdtlwawekfyldkqhapltleikdrq-----lvvllrredvlgqr 75  
 QY 73 LSSGOMFPQSOLPAVMQLYPGRKRYRADSPWEIADHGQIDSEOLVLYTPE 125  
 Db 76 pmtcpqjpslpllmwqlypdgryssdsstwrlyvnhikidgedvmllellpd 128  
 Db  
 RESULT 6  
 AAR94975  
 ID AAR94975 standard; Protein; 113 AA.  
 AC  
 XX AAR94975;  
 AC  
 XX 23-AUG-1996 (first entry)  
 DT  
 XX Human TCL-1 polypeptide.  
 DE  
 XX TCL-1; chromosome-14; leukaemia; lymphoma; therapy.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 31  
 FT Misc-difference 46 /note= "unidentified amino acid"  
 FT Misc-difference 46 /note= "unidentified amino acid"  
 FT Misc-difference 47 /note= "unidentified amino acid"  
 FT Misc-difference 55 /note= "unidentified amino acid"  
 FT Misc-difference 55 /note= "unidentified amino acid"  
 FT  
 XX W09613514-A1.  
 PN  
 XX 09-MAY-1996.  
 PD  
 XX 23-OCT-1995; 95WO-US13663.  
 PF  
 XX 27-OCT-1994; 94US-0330272.  
 PR  
 XX (RAGG-) RAGGIO-ITALGENE SPA.  
 PA

PA (UYJE-) UNIV JEFFERSON THOMAS.  
 XX  
 XX Croce CM, Russo G;  
 PI  
 XX WPI: 1996-239444/24.  
 DR  
 XX N-PSDB; AAT18877.  
 DR  
 XX  
 XX New TCL-1 protein and gene associated with chromosome 14  
 PT abnormalities - useful to develop prods. for detection, treatment  
 PT and prevention of diseases such as T-cell leukaemia(s) and  
 PT lymphoma(s)  
 PS  
 CC Disclosure; Page 69-72; 105pp; English.  
 CC  
 CC The amino acid sequence (AAR94975) of human TCL-1 protein was  
 CC deduced from the exon sequences of the TCL-1 gene (AAT18877).  
 CC A sequence deduced from a cDNA clone is given in AAR94974.  
 CC The TCL-1 gene is expressed at high levels in leukaemic cells  
 CC carrying a t(14;14)(q11;q32) translocation or an inv(14)(q11;q32)  
 CC inversion. The TCL-1 protein, and antibodies raised against it,  
 CC can be used for the diagnosis or treatment of conditions associated  
 CC with increased expression of TCL-1 proteins and/or with chromosome  
 CC 14 abnormalities, esp. T-cell leukaemia and lymphoma.  
 CC  
 XX  
 XX Sequence 113 AA:  
 SQ  
 Query Match 19.3%; Score 132; DB 17; Length 113;  
 Best Local Similarity 26.8%; Pred. No. 7, 6e-08;  
 Matches 33; Conservative 17; Mismatches 39; Indels 34; Gaps 3;  
 QY 13 PGRMTIORPGIYDEDEGRVTWTVVVRNPSSRRMARASQGSRYESTIVH 62  
 Db 14 pdrlwawekfyldkqhapltleikdrqxxvllrredvlgprmtqtxgpl--- 70  
 QY 63 LMQMAYHTRRELLSSGOMFPQSOLPAVMQLYPGRKRYRADSPWEIADHGQIDSEOLVLYT 122  
 Db 71 -----lpmlwqlypdgryssdsstwrlyvnhikidgedvmllell 109  
 QY 123 QPE 125  
 Db 110 lpd 112  
 Db  
 RESULT 7  
 AAE05866  
 ID AAE05866 standard; Protein; 513 AA.  
 AC  
 XX AAE05866;  
 AC  
 XX 24-SEP-2001 (first entry)  
 DT  
 XX Pseudomonas stutzeri open reading frame-O (ORF-O) protein.  
 DE  
 XX Pseudomonas stutzeri  
 KW  
 XX Open reading frame-O; ORF-O; pyridine-2,6-bis (thiocarboxylate);  
 KW PUTC; environmental remediation; phytoremediation; bioaccumulation;  
 KW water purification; solution mining mobilisation; immobilisation;  
 KW detoxification; redox state modifier; metal ion reactivity;  
 KW carbon tetrachloride; metal.  
 KW  
 XX Pseudomonas stutzeri.  
 OS  
 XX  
 XX W0200153309-A1.  
 PN  
 XX 26-JUL-2001.  
 PD  
 XX 19-JAN-2001; 2001WO-US02386.  
 PF  
 XX 20-JAN-2000; 2000US-0177251.  
 PR  
 XX (IDAH-) IDAHO RES FOUND INC.  
 PA (PASZ/) PASZCZYNSKI A.  
 PA (SEBA/) SEBAT J L.  
 PA

PT	branching enzyme sequences		
XX			
PS	Disclosure; Fig 8: 108pp; English.		
XX			
CC	The present sequence represents a potato starch branching enzyme (SBE).	10.2%	Score 70; DB 19; Length 906.
CC	The specification describes a method for affecting enzymatic activity in		
CC	a plant. The method comprises expressing in the plant a nucleotide		
CC	sequence encoding, partially or completely, an intron of a class A potato		
CC	SBE in a sense orientation, optionally together with a nucleotide		
CC	sequence which codes, partially or completely, for an intron of a class B		
CC	SBE in a sense or antisense orientation, and where the nucleotide		
CC	sequence does not contain a sequence that is sense to an exon sequence		
CC	normally associated with the intron. The method can be used for providing		
CC	genetically modified plants which are capable of producing modified and		
CC	improved starches whose properties would satisfy various industrial		
CC	requirements. They can be used for preparing tailor-made starches in		
CC	plants which could replace the post-harvest modified starches. They can		
CC	also be used for expressing genes of interest, e.g. pharmaceutically		
CC	active proteins or enzymes or proteins which are beneficial to plants.		
XX			
SQ	Sequence 906 AA;		
Query Match			
Best Local Similarity		24.2%	Posit. No. 24.

Query Match	10.2%;	Score 70;	DB 19;	Length 906;
Best Local Similarity	24.2%;	Pred. No. 24;		
Matches 31;	Conservative 17;	Mismatches 34;	Indels 46;	Gaps 8
QY	17	WIOHPGIDEDEGRTWTVVVRFPNPSREEMRA-SQGRYERSTVHLMQMAVPTRE---	72	
Db	621	wldpf-----rgnmw-----sydkerrgmlaselltyk---lmaafatamsidekf	667	
QY	73	-LSSGQMPFS-----QDPANWQIYPGR-----KYRAA-DSGF	103	
Db	668	sflasgkqivssmdddnkvvferglvfyfnfphknteygikvgcdlpkkyvaldsda	727	
QY	104	WEIADHGO_111		
Db	728	wefgghgr_735		

	RESULT	9
AAW69300	ID	AAW69300 standard; Protein; 906 AA.
XX	AC	AAW69300;
XX	DT	07-DEC-1998 (first entry)
XX	DE	Potato class B starch branching enzyme.
XX	KW	Starch branching enzyme; SBE; potato; antisense; amylopectin; transgenic plant.
XX	SOS	Solanum tuberosum cv. Desiree.
XX	PN	M09837213-A1.
XX	PD	27-AUG-1998.
XX	PP	23-FEB-1996; 98WO-IB00270.
XX	PR	24-MAR-1997; 97GB-0006060.
XX	PA	21-FEB-1997; 97GB-0003663.
XX	(DANT-)	DANISCO AS.
XX	Poulsen P;	
XX	WPI:	1998-467573/40.
XX	N-PDB:	AAV55069.
XX		affecting enzymatic activity using antisense intron inhibition - especially of starch branching enzyme in plants, useful to alter

PT starch branching enzyme activity, amylopectin levels or starch composition

XX Example: Fig 12; 95pp; English.

CC This is the amino acid sequence of the class B starch branching enzyme (SBE) of potato cv. Desiree. It was deduced from a full-length SBE gene sequence (see AA55069). A novel method of affecting enzymatic activity in plants or other starch-producing organisms (or their cells, tissues or organs) comprises expressing a nucleotide sequence coding for an intron (see AA55046) of a class A potato SBE in an antisense orientation, but not containing a sequence antisense to an exon sequence normally associated with the intron. The sequence is optionally expressed with a nucleotide sequence which codes, partially or completely, for an intron of a class B SBE in an antisense/sense orientation. Also claimed are a promoter (see AA55047) useful in expressing the nucleotide sequence of the method, a transgenic starch-producing organism, and the starch produced by the organism. The method and sequences enable production of transgenic starch-producing organisms with altered characteristics, and particularly plants with altered SBE activity, amylopectin levels and/or starch composition.

CC Sequence 906 AA:

Query Match 10.2%; Score 70; DB 19; Length 906;

Best Local Similarity 24.2%, Pred. No. 24; Matches 31; Conservative 17; Mismatches 34; Indels 46; Gaps 8;

QY 17 WIDRPGIYDEGRWTVVYVRFNPSRREWAR-SQGSRYEPTVHLMQAVHRE--- 72  
DB 621 WIDF-----regnw-----sydkcrqvwlnadshlyk---fmafdramusldexf 667  
QY 73 -LSSGQMFPS-----QLPAVWQLYPR-----KYRAA-DSSP 103  
DB 668 sfiasgkqivsgmddnkvvrfvgrglvfvfnfhpkhtygykvvgcdlpgyrvvaldsda 727  
QY 104 WETADHGQ 111  
DB 728 wefgghyr 735

RESULT 10

AAW14740 ID AAW14740 standard; Protein: 476 AA.

XX AAW14740:

XX 12-MAY-1997 (first entry)

XX Protein encoded by IBDV clone GLS-3.

XX Infectious bursal disease virus; IBDV; large segment; GLS 5 strain;

XX structural protein; VP2; VP3; VP4; Immunogen; vaccine;

XX Gumboro disease.

XX Infectious bursal disease virus.

XX US5595912-A.

XX 21-JAN-1997.

XX 04-MAY-1990; 90US-0519202.

XX 23-MAR-1994; 94US-0216276.

XX 04-MAY-1990; 90US-0519202.

XX 28-JUN-1993; 93US-0083784.

XX (UYMA-) UNITV MARYLAND BALTIMORE.

XX Snyder D, Vakharlia V;

XX PI

DR WPI: 1997-107584/10.

DR N-PSDB; AAT63068.

XX Infectious bursal disease virus nucleic acids - for prodn. of

XX vaccines for poultry for protection against Gumboro disease

XX Claim 5; Column 71-76; 41pp; English.

CC The sequences given in AAW14738-41 are encoded by fragments of the infectious bursal disease virus (IBDV) large segment from the GLS 5 strain. These sequences are encoded by clones GLS-1 to GLS-4. Proteins such as these, derived from IBDV, esp. VP2 which is the major host protective immunogen of IBDV, may be used to prepare vaccines against IBDV infection/Gumboro disease.

CC Sequence 476 AA:

Query Match 10.1%; Score 69; DB 18; Length 476;

Best Local Similarity 23.8%, Pred. No. 13; Matches 30; Conservative 18; Mismatches 50; Indels 28; Gaps 5;

QY 8 RIGVPRGLMTQRPGEYDEGRWTVVYVRFNPSRREWAR-----ASGSRYP 57  
DB 162 rlg-----lkaqpgatdvcnqpnwafikrfrprowdrlpylnlpylppnagry-- 214  
QY 58 SITVHLMQMAVHREU--LSSGQMFSQLPAVWQLYPGKRYRAADSSFEIADHQCIDSM 115  
DB 215 ----lhamasefkehpelasesavrameaasvdp1-----lqsalvfmwleengivtdm 265  
QY 116 EQLVLT 121  
DB 266 anfaIs 271

RESULT 11

AAW14741 ID AAW14741 standard; Protein: 540 AA.

XX AAW14741:

XX 12-MAY-1997 (first entry)

XX Protein encoded by IBDV clone GLS-4.

XX Infectious bursal disease virus; IBDV; large segment; GLS 5 strain;

XX structural protein; VP2; VP3; VP4; immunogen; vaccine;

XX Gumboro disease.

XX Infectious bursal disease virus.

XX US5595912-A.

XX 21-JAN-1997.

XX 04-MAY-1990; 90US-0519202.

XX 23-MAR-1994; 94US-0216276.

XX 04-MAY-1990; 90US-0519202.

XX 28-JUN-1993; 93US-0083784.

XX (UYMA-) UNITV MARYLAND BALTIMORE.

XX Snyder D, Vakharlia V;

XX WPI: 1997-107584/10.

XX N-PSDB; AAT63069.

XX Infectious bursal disease virus nucleic acids - for prodn. of

XX vaccines for poultry for protection against Gumboro disease

XX Claim 2; Column 65-70; 41pp; English.



Db 751 ----hlmaseefkelpesavrameaasvdp1-----fgsalsvfmwleenglytdm 801  
 Oy 116 EOLVLT 121  
 Db 802 anfalls 807

## RESULT 14

AAPE1316  
 ID AAPE1316 standard; Protein: 1021 AA.

AC AAPE1316;

DT 29-OCT-1991 (first entry)

DE Infectious bursal disease virus 32KD structural protein.

KW IBDV; poultry; vaccine.

OS Avian Infectious bursal disease virus.

PN WO8607060-A.

PD 04-DEC-1986.

PE 30-MAY-1986; 86WO-AU00156.

PF 23-AUG-1985; 85AU-0002118.

PR 30-MAY-1985; 85AU-0000815.

PT 01-JAN-1986; 86AU-0039668.

PA (CSIR ) COMMONWEALTH SCIENT ORG.

PI Azad AA, Hudson PJ, Fahey KJ;

DR WPI: 1986-332075/50.

DR N-PSDB; AAN60874.

PF Recombinant DNA corresp. to infectious bursal disease virus-RNA -

PT used for producing polypeptide(s) for use in treatment of

PT Infectious bursal disease virus in chickens.

PS Claim 10; Fig 10; 78pp; English.

XX The avian IBDV 32KD structural protein product may be used in

XX serological diagnosis and development of a vaccine against the

XX disease. The protein is encoded by the only extensive reading frame

CC of the sequence.

CC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.

XX The sequence is that encoded by variant GLS clones 1 to 4 of infectious

XX bursal disease virus (IBDV). It can be used in broad spectrum IBDV

XX poultry vaccines, administration is optimally, by injection,

XX nasally or orally, at any time after hatching and as a booster at

XX other times. Poultry treated include chickens, roosters, broilers,

XX roasters, breeders, layers, turkeys and ducks. See also AARI5053.

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.

XX The sequence is that encoded by variant GLS clones 1 to 4 of infectious

XX bursal disease virus (IBDV). It can be used in broad spectrum IBDV

XX poultry vaccines, administration is optimally, by injection,

XX nasally or orally, at any time after hatching and as a booster at

XX other times. Poultry treated include chickens, roosters, broilers,

XX roasters, breeders, layers, turkeys and ducks. See also AARI5053.

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.

XX The sequence is that encoded by variant GLS clones 1 to 4 of infectious

XX bursal disease virus (IBDV). It can be used in broad spectrum IBDV

XX poultry vaccines, administration is optimally, by injection,

XX nasally or orally, at any time after hatching and as a booster at

XX other times. Poultry treated include chickens, roosters, broilers,

XX roasters, breeders, layers, turkeys and ducks. See also AARI5053.

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.

XX The sequence is that encoded by variant GLS clones 1 to 4 of infectious

XX bursal disease virus (IBDV). It can be used in broad spectrum IBDV

XX poultry vaccines, administration is optimally, by injection,

XX nasally or orally, at any time after hatching and as a booster at

XX other times. Poultry treated include chickens, roosters, broilers,

XX roasters, breeders, layers, turkeys and ducks. See also AARI5053.

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.

XX The sequence is that encoded by variant GLS clones 1 to 4 of infectious

XX bursal disease virus (IBDV). It can be used in broad spectrum IBDV

XX poultry vaccines, administration is optimally, by injection,

XX nasally or orally, at any time after hatching and as a booster at

XX other times. Poultry treated include chickens, roosters, broilers,

XX roasters, breeders, layers, turkeys and ducks. See also AARI5053.

XX

XX

XX

XX

XX AARI5054;  
 AC  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Polypeptide encoded GLS clones 1 to 4.  
 XX  
 KW Infectious bursal disease virus; IBDV; vaccine; poultry.  
 XX  
 OS Infectious bursal disease virus.

XX Key

XX Location/Qualifiers

FT 1..348

FT /tag= b

FT /note= "GLS-1 clone"

FT 283..1252

FT /tag= c

FT /note= "GLS-2 clone"

FT 1722..3230

FT /tag= d

FT /note= "GLS-3 clone"

FT 999..2620

FT /tag= e

FT /note= "GLS-4 clone"

XX WO9116925-A.

XX 14-NOV-1991.

XX 30-APR-1991; 91WO-US03056.

XX 04-MAY-1990; 90US-0519202.

XX (UYMA-) UNIV MARYLAND COL.

XX Vakharia V;

XX WPI: 1991-353531/48.

XX N-PSDB; AAQ14897.

XX Infectious bursal disease virus nucleic acids - useful as

XX vaccines against IBDV in poultry

XX

XX Example: Page 30; 45pp; English.



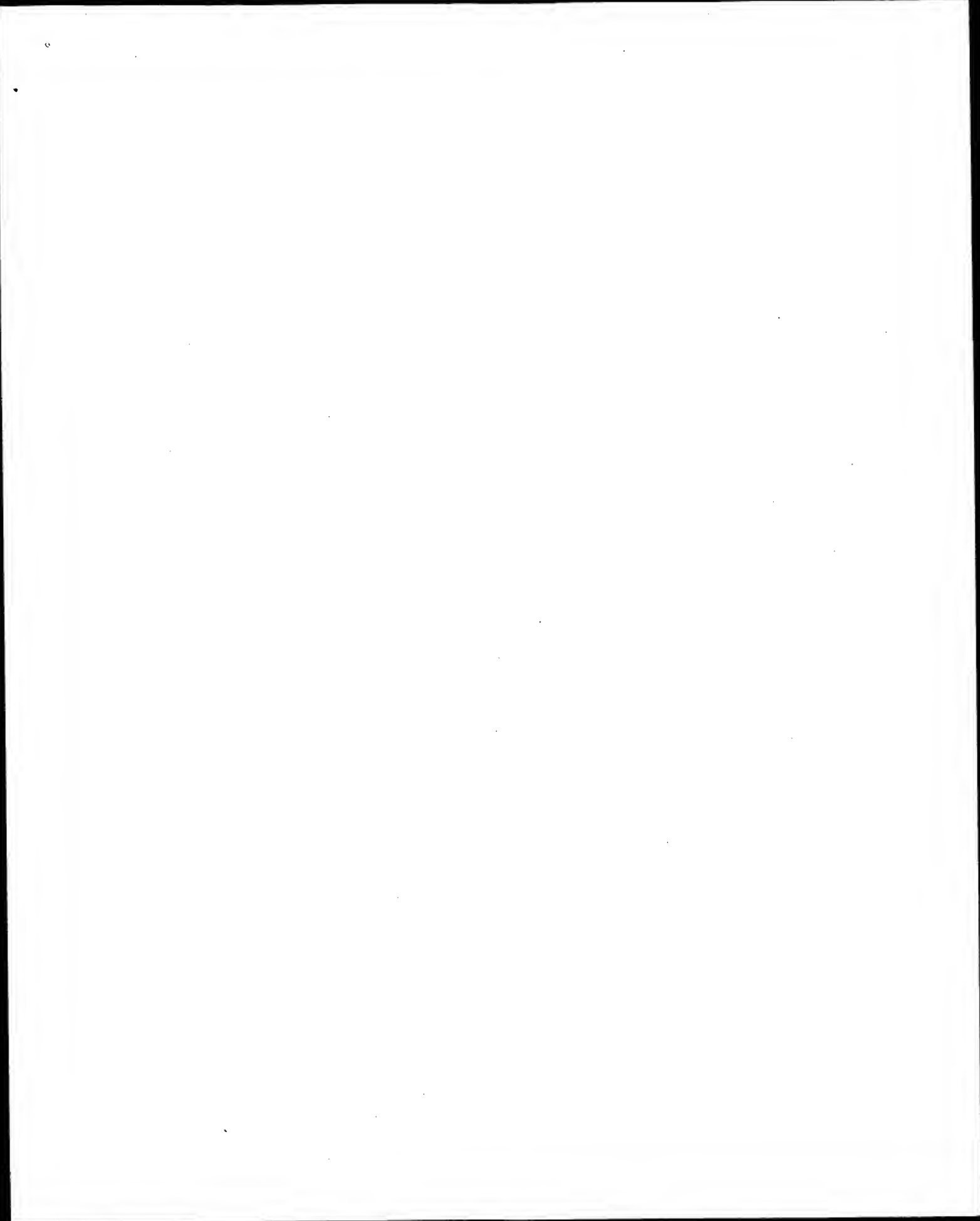
•  
Fri Nov 30 11:03:09 2001

us-09-526-329-39.rag

Page 9

Search completed: November 29, 2001, 04:03:10  
Job time: 516 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2001, 03:58:04 ; Search time 14.71 Seconds  
(without alignments)  
195.814 Million cell updates/sec

Title: US-09-526-329-39

Perfect score: 685

Sequence: 1 MASEASVRLGVPGRILWICR.....HGQIDSMEDVLVTPERKD 128

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents: AA: \*  
1: /cgn2\_6/ptodata/2/laa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/laa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/laa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/laa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/laa/PCUTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/laa/beckfilest1.pep: \*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176.5	25.8	108	2	US-08-330-272-4
2	176.5	25.8	108	5	PCT-US95-13663-4
3	149	21.8	113	5	US-08-330-272-2
4	149	21.8	113	5	PCT-US95-13663-2
5	69	10.1	476	1	US-08-216-276A-33
6	69	10.1	540	1	US-08-216-276A-31
7	69	10.1	1012	1	US-08-216-276A-19
8	69	10.1	1012	1	US-08-219-262B-1
9	69	10.1	1012	1	US-08-219-262B-9
10	69	10.1	1012	1	US-08-219-262B-12
11	69	10.1	1012	3	US-09-031-655-1
12	69	10.1	1012	3	US-09-031-655-9
13	69	10.1	1012	3	US-09-031-655-12
14	68	9.9	1012	2	US-08-708-541A-34
15	67.5	9.9	1012	2	US-08-219-262B-7
16	67.5	9.9	1012	4	US-09-031-655-7
17	67.5	9.9	3072	4	US-09-031-655-3
18	67.5	9.9	3079	4	US-09-031-655-3
19	66	9.6	1012	1	US-07-944-525-2
20	66	9.6	1012	1	US-08-219-262B-3
21	66	9.6	1012	1	US-08-219-262B-5
22	66	9.6	1012	1	US-08-219-262B-6
23	66	9.6	1012	1	US-08-219-262B-8
24	66	9.6	1012	2	US-08-708-541A-30
25	66	9.6	1012	2	US-09-031-655-3
26	66	9.6	1012	3	US-09-031-655-3
27	66	9.6	1012	3	US-09-031-655-3

28	66	9.6	1012	3	US-09-031-655-5	Sequence 5, Appl1
29	66	9.6	1012	3	US-09-031-655-6	Sequence 6, Appl1
30	66	9.6	1012	3	US-09-031-655-8	Sequence 8, Appl1
31	65	9.5	276	2	US-08-712-072C-4	Sequence 4, Appl1
32	65	9.5	797	2	US-08-663-566A-2	Sequence 2, Appl1
33	65	9.5	797	2	US-08-023-610-2	Sequence 2, Appl1
34	65	9.5	797	2	US-08-288-065A-2	Sequence 2, Appl1
35	65	9.5	797	2	US-08-362-240B-2	Sequence 2, Appl1
36	65	9.5	797	5	PCT-US95-10245-2	Sequence 2, Appl1
37	65	9.5	1012	1	US-07-944-943-2	Sequence 2, Appl1
38	65	9.5	1012	1	US-08-219-262B-2	Sequence 2, Appl1
39	65	9.5	1012	1	US-08-219-262B-14	Sequence 2, Appl1
40	65	9.5	1012	3	US-09-031-655-2	Sequence 14, Appl1
41	65	9.5	1012	3	US-09-031-655-14	Sequence 14, Appl1
42	64.5	9.4	332	4	US-09-331-581-23	Sequence 23, Appl1
43	64	9.3	380	1	US-08-416-478A-6	Sequence 6, Appl1
44	64	9.3	380	2	US-08-474-988B-6	Sequence 6, Appl1
45	64	9.3	380	2	US-08-394-442B-6	Sequence 6, Appl1

## ALIGNMENTS

```

RESULT 1
US-08-330-272-4
Sequence 4, Application US/08330272
Patent No. 5985598
GENERAL INFORMATION:
APPLICANT: Russo et al
TITLE OF INVENTION: TCI-1 Gene and Protein and Related
NUMBER OF INVENTIONS: 11
CORRESPONDENCE ADDRESSES:
ADDRESS: Pennine & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,272
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrick, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6754-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 790-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-330-272-4

```

Query Match 25.8%; Score 176.5; DB 2; Length 108;  
Best Local Similarity 33.6%; Pred. No. 6.5e-14;  
Matches 43; Conservative 17; Mismatches 43; Indels 25; Gaps 4;  
QY 1 MASEASVRLGVPGRILWICR...-VFNPSSRRMARAGSGSYEP 57  
DB 1 MAGE---DVGAPPHLWVHGEGYRDEYGRITVAVVEERSFLIRARVOIQVPLGDARP 57

OY 58 SITVHLQMAVHTRELLSSGQMPISQPAVWQLYPGKRYRADSSFWELADHGOIDSMEO 117  
 DB 58 S-----HLLT-----SOLPLMWOLYPERRYMDNNSRLMOIGHHLWVGVQE 98  
 OY 118 LVLTQPE 125  
 DB 99 LILKLPLD 106

RESULT 2  
 PCT-US95-13663-4  
 ; Sequence 4, Application PC/TUS9513663  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Russo et al  
 ; TITLE OF INVENTION: TCL-1 Gene and Protein and Related  
 ; TITLE OF INVENTION: Methods and Compositions  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennile & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/13663  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 790-8864/9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 108 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; PCT-US95-13663-4

Query Match 25.8%; Score 176.5; DB 5; Length 108;  
 Best Local Similarity 33.6%; Pred. No. 6,5e-14;  
 Matches 43; Conservative 17; Mismatches 43; Indels 25; Gaps 4;  
 OY 1 MASEAVRLGVPGRTIOTRPGTYDESGRTWTVV---VRFNPSRREMARASOGSHYEP 57  
 DB 1 MAGE---DYGAPDHLWROEGITREIVRTVAVVEELSFARAVQIDVPLGDARP 57  
 OY 58 SITVHLQMAVHTRELLSSGQMPISQPAVWQLYPGKRYRADSSFWELADHGOIDSMEO 117  
 DB 58 S-----HLLT-----SOLPLMWOLYPERRYMDNNSRLMOIGHHLWVGVQE 98  
 OY 118 LVLTQPE 125  
 DB 99 LILKLPLD 106

RESULT 3  
 US-08-330-272-2  
 ; Sequence 2, Application US/08330272  
 ; Patent No. 5985598

GENERAL INFORMATION:  
 ; APPLICANT: Russo et al  
 ; TITLE OF INVENTION: TCL-1 Gene and Protein and Related  
 ; TITLE OF INVENTION: Methods and Compositions  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennile & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/330,272  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 790-8864/9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 113 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-330-272-2

Query Match 21.8%; Score 149; DB 2; Length 113;  
 Best Local Similarity 29.2%; Pred. No. 1.4e-10;  
 Matches 33; Conservative 21; Mismatches 45; Indels 14; Gaps 1;  
 OY 13 PGLTWIORPGTYDESGRTWTVV---VRFNPSRREMARASOGSHYEPSTVHLQMAVHTRE 72  
 DB 14 PDLWAMEKRYVIDERQHMMLTETIKDRDQ-----LWLLRREDVLCGR 59  
 OY 73 LSSGQMPISQPAVWQLYPGKRYRADSSFWELADHGOIDSMEOIVLTQPE 125  
 DB 60 PMTPQIGPSLPLIMWOLYPERYSDDSSFWRLVTHIKIDGVEDMLLELPLD 112

RESULT 4  
 PCT-US95-13663-2  
 ; Sequence 2, Application PC/TUS9513663  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Russo et al  
 ; TITLE OF INVENTION: TCL-1 Gene and Protein and Related  
 ; TITLE OF INVENTION: Methods and Compositions  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennile & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/13663

```

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6754-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 790-8864/9741
; TELETYPE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-13663-2

```

```

Query Match      21.8%; Score 149; DB 5; Length 113;
Best Local Similarity 29.2%; Pred. No. 1.4e-10;
Matches 33; Conservative 21; Mismatches 45; Indels 14; Gaps 1;

OY 13 PGRMTIORGIYEDDEGRVTWVYVFNPSRREMARASQGSRYESTTYHLMQAVHTRE 72
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 14 PDLTAMEKFFYLDEKOHANLPLTITIKRQ-----LRYLLRRDDVYVLR 59
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 73 LSSGOMPSPQLPAVMQLYPGRKRYAADSFWETADHGQIDSMQVLTPOPE 125
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 60 PMPTQIGSLPLIMQLYPDGRYSSDSFWRLVYH1KIDGVEDMLLELPLD 112
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

RESULT 5
US-08-216-276A-33
; Sequence 33, Application US/08216276A
; Patent No. 5595912
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, VIKRAM
; APPLICANT: SNYDER, DAVID
; TITLE OF INVENTION: SPECIFIC DNA AND RNA SEQUENCES
; TITLE OF INVENTION: ASSOCIATED WITH US 1BDV VARIANTS, VECTOR CARRYING DNA
; TITLE OF INVENTION: SEQUENCES, HOST CARRYING CLONED VECTOR, DEDUCED AMINO ACID
; TITLE OF INVENTION: SEQUENCES, VACCINE AND METHOD OF VACCINATION
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,276A
; FILING DATE: 23-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/083,784
; FILING DATE: 28-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/519,202
; FILING DATE: 04-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/227,311
; FILING DATE: 02-AUG-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Kelber, Steven B.
; REGISTRATION NUMBER: 30,073

```

```

; REFERENCE/DOCKET NUMBER: 2747-054-27 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELETYPE: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-276A-33

```

```

Query Match      10.1%; Score 69; DB 1; Length 476;
Best Local Similarity 23.8%; Pred. No. 3.8;
Matches 30; Conservative 18; Mismatches 50; Indels 28; Gaps 5;

OY 8 RLGVPPGRMTIORGIYEDDEGRVTWVYVFNPSRREMAR-----ASQGSRYEP 57
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 162 RLG-----LKLGGGADVNTGPMATFKRFPHPMDRLPYLPPNAGROY-- 214
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 58 SITVHMQAVHTREL--LSSGOMPSPQLPAVMQLYPGRKRYAADSFWETADHGQIDSM 115
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 215 ---HLMAASGEPPEPELESAYRAMEAASVDPL-----FQSALSYVWMLLENGIVTDM 265
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 116 EQVLYLT 121
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 266 ANPALS 271
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

RESULT 6
US-08-216-276A-31
; Sequence 31, Application US/08216276A
; Patent No. 5595912
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, VIKRAM
; APPLICANT: SNYDER, DAVID
; TITLE OF INVENTION: SPECIFIC DNA AND RNA SEQUENCES
; TITLE OF INVENTION: ASSOCIATED WITH US 1BDV VARIANTS, VECTOR CARRYING DNA
; TITLE OF INVENTION: SEQUENCES, HOST CARRYING CLONED VECTOR, DEDUCED AMINO ACID
; TITLE OF INVENTION: SEQUENCES, VACCINE AND METHOD OF VACCINATION
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,276A
; FILING DATE: 23-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/083,784
; FILING DATE: 28-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/519,202
; FILING DATE: 04-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/227,311
; FILING DATE: 02-AUG-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Kelber, Steven B.
; REGISTRATION NUMBER: 30,073
; REFERENCE/DOCKET NUMBER: 2747-054-27 CIP

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 540 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-216-276A-31

Query Match 10.1%; Score 69; DB 1; Length 540;  
Best Local Similarity 23.8%; Pred. No. 4.5;  
Matches 30; Conservative 18; Mismatches 50; Indels 28; Gaps 5;

QY 8 RLGVPGLMIOBPGIYEDDEGRVTWVYVRFNPSSREMAR-----ASQGSRYEP 57  
DB 403 RLG-----LKLGGAFDVTGPNWATFIRFPHNRDRLPYLNLPLPNAHQY-- 455  
QY 58 SITVHLQMAVHTREL--LSSQMPFSQLPVWQLYPGKRYRADSSFEIADHQIDSM 115  
DB 456 ----HMAAASEKETPELESVVRAMEAASVDPL-----FQSALSVFMLENGIVTDM 506  
QY 116 EQVLVT 121  
DB 507 ANFALS 512

RESULT 7  
US-08-216-276A-19  
Sequence 19, Application US/08216276A  
Patent No. 5595912  
GENERAL INFORMATION:  
APPLICANT: VAKHARIA, VIKRAM  
TITLE OF INVENTION: SPECIFIC DNA AND RNA SEQUENCES  
TITLE OF INVENTION: ASSOCIATED WITH US 187V VARIANTS, VECTOR CARRYING DNA  
TITLE OF INVENTION: SEQUENCES, HOST CARRYING CLONED VECTOR, DEDICED AMINO ACID  
TITLE OF INVENTION: SEQUENCES, VACCINE AND METHOD OF VACCINATION  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,276A  
FILING DATE: 23-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US 08/083,784  
FILING DATE: 28-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/519,202  
FILING DATE: 04-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/227,311  
FILING DATE: 02-AUG-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Kelder, Steven B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-054-27 CIP  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1012 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-216-276A-19

Query Match 10.1%; Score 69; DB 1; Length 1012;  
Best Local Similarity 23.8%; Pred. No. 11;  
Matches 30; Conservative 18; Mismatches 50; Indels 28; Gaps 5;

QY 8 RLGVPGLMIOBPGIYEDDEGRVTWVYVRFNPSSREMAR-----ASQGSRYEP 57  
DB 698 RLG-----LKLGGAFDVTGPNWATFIRFPHNRDRLPYLNLPLPNAHQY-- 750  
QY 58 SITVHLQMAVHTREL--LSSQMPFSQLPVWQLYPGKRYRADSSFEIADHQIDSM 115  
DB 751 ----HMAAASEKETPELESVVRAMEAASVDPL-----FQSALSVFMLENGIVTDM 801  
QY 116 EQVLVT 121  
DB 802 ANFALS 807

RESULT 8  
US-08-219-262B-1  
Sequence 1, Application US/08219262B  
Patent No. 5788970  
GENERAL INFORMATION:  
APPLICANT: VAKHARIA, VIKRAM  
TITLE OF INVENTION: MENDEL-WHERSAT, STEPHANIE A  
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS  
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED  
TITLE OF INVENTION: THEREON  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT  
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/219,262B  
FILING DATE: 29-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2747-047-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1012 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:







## GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM  
 APPLICANT: SNYDER, DAVID B  
 APPLICANT: MENGEL-WHERSAT, STEPHANIE A  
 TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS  
 TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED  
 TITLE OF INVENTION: THEREON  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
 STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: USA  
 ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/031,655  
 FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/219,262

FILING DATE: 29-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: OBION, NORMAN F

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2747-047-27

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1012 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-031-655-12

## Query Match

Best Local Similarity 10.1%; Score 69; DB 3; Length 1012;

Matches 30; Conservative 18; Mismatches 50; Indels 28; Gaps 5;

QY 8 RLGVPGRLMIORPGIYEDSGRTVTVVRFNPSRREMAR-----ASGSGREP 57

DB 698 RLGVPGRLMIORPGIYEDSGRTVTVVRFNPSRREMAR-----ASGSGREP 57

QY 58 STVHLMQMVHTREL--LSSGMPFSQLPAYWQLYPGKRYRADSSFWELADHGQIDSM 115

DB 751 ----HMAASERKETPELESVVRAMEAASVDPL-----FQSALSVFWMLENGIYIDM 801

QY 116 RQLVLT 121

DB 802 ANFALS 807

## RESULT 14

US-08-708-541A-34

; Sequence 34, Application US/08708541A

; Patent No. 5871744

; GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM N.

APPLICANT: MONDT, Egbert

TITLE OF INVENTION: A METHOD FOR GENERATING BIRNAVIRUS FROM

TITLE OF INVENTION: SYNTHETIC RNA TRANSCRIPTS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP

STREET: 655 Fifteenth Street, N. W.,

STREET: Suite 330 - G Street Lobby

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/708,541A

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: KITTS, Monica C.

REGISTRATION NUMBER: 36,105

REFERENCE/DOCKET NUMBER: P8172-6002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202/638-4810

TELEFAX: 202/638-4810

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 1013 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-708-541A-34

## Query Match

Best Local Similarity 9.9%; Score 68; DB 2; Length 1013;

Matches 27; Conservative 19; Mismatches 61; Indels 10; Gaps 4;

QY 8 RLGVPGRLMIORPGIYEDSGRTVTVVRFNPSRREMARASQ--GSRVPSSTVTHQM 66

DB 699 RLGMK-----LAGGAYDINTGPMNATFVRFRPHNFRMDRLPYLNLPIPTPAGROPHL 753

QY 67 AVHTRELLSSGOM--FQSOLPAYWQLYPGKRYRADSSFWELADHGQIDSMQVLT 121

DB 754 ALMAASERKETPELEDAVRAMDAANADP--LFRSALOYFWMLENGIYIDMANFALS 808

## RESULT 15

US-08-219-262B-7

; Sequence 7, Application US/08219262B

; Patent No. 5788970

; GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM

APPLICANT: SNYDER, DAVID B

TITLE OF INVENTION: MENGEL-WHERSAT, STEPHANIE A

TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS

TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/219,262B

FILING DATE: 29-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F  
 REGISTRATION NUMBER: 24, 618  
 REFERENCE/DOCKET NUMBER: 2747-047-27  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 413-3000  
 TELEFAX: (703) 413-2220  
 TELETYPE: 248855 OPAT UR  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1012 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOTIF TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Infectious bursal disease virus  
 STRAIN: 52/70  
 US-08-219-262B-7

Query Match 9.9%; Score 67.5; DB 1; Length 1012;  
 Best local similarity 23.9%; Pred. No. 16;  
 Matches 26; Conservative 14; Mismatches 28; Indels 41; Gaps 5;  
 OY 8 RIGVPPGRLMIQRPQIYDEBGRVTYVYVFNPSRRERARASQSKTEPSITVHLMQMA 67  
 DB 698 RLG-----LKLGGAGADVNTGSPNNATFIKREPHNPMDR----- 733  
 OY 68 VHTRELLSSGOMPFSQLPVWOLYP--GRKYRAADSSFEWELADHGQIDS 114  
 DB 734 -----LPYLNLP--YLPNAGRGYHLMAAS--EFKDTPELES 767

Search completed: November 29, 2001, 04:03:37  
 Job time: 333 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2001, 03:59:59 ; Search time 18.53 Seconds

(without alignments)  
526.192 Million cell updates/sec

Title: US-09-526-329-39

Perfect score: 685  
Sequence: 1 MASBASVRLGVPPGRLMTQR.....HQIDSMQVLVTPQPERKD 128

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Lasting first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176.5	25.8	107	2	MTCP-1 protein sp1
2	149	21.8	114	2	T cell leukemia/ly
3	79.5	11.6	447	2	embryonic muscle m
4	79.5	11.6	447	2	embryonic muscle m
5	79.5	11.6	603	2	dopamine beta-mono
6	79	11.5	238	2	probable transcrip
7	75.5	11.0	476	2	two-component sens
8	74.5	10.9	436	2	viid2 protein - Ag
9	73.5	10.7	261	2	hypothetical 30.1K
10	73	10.7	261	2	hypothetical 30.1K
11	72.5	10.6	463	2	probable metallo-o
12	72.5	10.6	424	2	hypothetical viid2
13	72.5	10.6	513	2	probable iron-sulf
14	70	10.2	1414	2	probable mbtB prot
15	70	10.2	367	2	N-ethylmaleimide r
16	70	10.2	830	2	1,4-alpha-glucan b
17	69.5	10.1	861	1	SHL protein - hum
18	69.5	10.1	357	2	hypothetical prote
19	69.5	10.1	363	2	hypothetical prote
20	69.5	10.1	1927	2	embryonic muscle m
21	69.5	10.1	363	2	conserved hypotet
22	69	10.1	1012	1	genome polyprotein
23	69	10.1	1012	1	vitellogenin - gyp
24	68	9.9	1013	2	hypothetical prote
25	67.5	9.9	1013	2	dolichyl-phosphate
26	67.5	9.9	1012	1	genome polyprotein
27	67.5	9.9	1148	2	hypothetical prote
28	67	9.8	125	2	succinate dehydrog
29	66.5	9.7	596	2	minor tail protein

30	66.5	9.7	1209	2	H85839	probable regulator
31	66	9.6	237	2	C83183	probable transcrip
32	66	9.6	363	2	T37072	hypothetical prote
33	66	9.6	536	2	D83622	arylsulfatase PA01
34	66	9.6	993	1	D83622	genome polyprotein
35	66	9.6	1001	2	C86181	hypothetical prote
36	66	9.6	1012	1	GNXSIR	genome polyprotein
37	66	9.6	1012	1	GNXSIR	genome polyprotein
38	66	9.6	1170	2	H71295	probable DNA polym
39	66	9.6	1175	2	T25634	hypothetical prote
40	65.5	9.6	327	2	T15594	hypothetical prote
41	65.5	9.6	350	2	D83102	probable receptor
42	65.5	9.6	764	2	D84847	probable receptor
43	65.5	9.5	286	2	S48201	lipoxygenase (EC 3.
44	65	9.5	352	2	S76078	hypothetical prote
45	65	9.5	406	2	C83867	Xaa-Pro dipeptidase

## ALIGNMENTS

RESULT 1  
S78532  
MTCP-1 protein splice form B1 - human  
C:Species: Homo sapiens (man)  
C>Date: 17-Jul-1998 #sequence\_rev15ion 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: S78532  
R:Stern, M.H.; Soulier, J.; Rosenzweig, M.; Nakahara, K.; Canki-Klajn, N.; Aurias, A.  
Oncogene 8, 2475-2483, 1993  
A>Title: MTCP-1: a novel gene on the human chromosome Xq28 translocated to the T cell  
A:Reference number: 138045; MUID: 93368950  
A:Accession: S78532  
A:Molecule type: DNA  
A:Residues: 1-107 <STE>  
A:Cross-references: EMBL:Z24459; NID:g2252491; PIDN:CAA80828.1; PID:g2252492  
C:Genetics:  
A:Gene: MTCP-1  
A:Introns: 35/3; 92/3  
C:Keywords: alternative splicing; T-cell proliferation

Query Match 25.8%; Score 176.5; DB 2; Length 107;  
Best Local Similarity 33.6%; Pred. No. 1.9e-11;  
Matches 43; Conservative 17; Mismatches 43; Indels 25; Gaps 4;  
QY 1 MASBASVRLGVPPGRLMTQRPGIYDEEGRTWTVV---VREMPSPREARASGSRTER 57  
DB 1 MAGE--DVGAPDHVWROGIRYRTVAVVAVEETSFIRARVQIOVPIGDAAAP 57  
QY 58 STVYHLMQNAVTRRELLSSGOMPFPQPAVMQVYPRKRYRAADSSFWETADHQIDSMQ 117  
DB 58 S-----HLLT-----SOLPLMQLYPERYMDNNSRLMOIHHLNVRGVE 98  
QY 118 LVLTQPE 125  
DB 99 LLKLKLPD 106  
RESULT 2  
T cell leukemia/lymphoma protein TCL1 - human  
N:Alternate names: T-cell leukemia-related protein TCL1  
C:Species: Homo sapiens (man)  
C>Date: 09-Mar-1996 #sequence\_rev15ion 09-Mar-1996 #text\_change 05-Nov-1999  
R:Virgilio, L.; Narducci, M.G.; Isobe, M.; Ball198, L.G.; Cooper, M.D.; Croce, C.M.;  
Proc. Natl. Acad. Sci. U.S.A. 91, 12530-12534, 1994  
A>Title: Identification of the TCL1 gene involved in T-cell malignancies.  
A:Reference number: 138286; MUID:95107991  
A:Accession: 138286  
A:Molecule type: mRNA  
A:Residues: 1-114 <RES>  
A:Cross-references: EMBL:X82240; NID:g624960; PIDN:CAA57708.1; PID:g624961

## C:Genetics:

A:Gene: GDB:TCL1  
A:Cross-references: GDB:250785  
A:Map position: 14q32.1-14q32.1

## Query Match

Best Local Similarity 21.8%; Score 149; DB 2; Length 114;  
Best Local Similarity 29.2%; Pred. No. 1.5e-08;  
Matches 33; Conservative 21; Mismatches 45; Indels 14; Gaps 1;

OY 13 PGLMTQREGTIDEGRWTVVVFNFNSRREWARASOGSRPESTIVYHLMQNAVHRE 72  
DB 15 PDLMTWMEFYVDEKQHWLPLITIKDRLO-----LRYLLRREDVVLGR 60  
OY 73 LSSGMPSPQAPVWOLYRGKRYRAADSPWEADIGQIDSMEQVYLTYOPE 125  
DB 61 PMPTQIGPSLLPIMWOLYPDGRYRSSDSEFWRLVYHRIKIDVEDMLLELPD 113

## RESULT 3

S11839  
vird2 protein - Agrobacterium tumefaciens plasmid pTIC58

C:Species: Agrobacterium tumefaciens  
C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 08-Oct-1999  
C:Accession: S11839; C22666; S12669  
R:Rogowsky, P.M.; Powell, B.S.; Shirsau, K.; Lin, T.S.; Morel, P.; Zyprian, E.M.; Steck, A.;  
Plasmid 23, 85-106, 1990  
A:Title: Molecular characterization of the vir regulon of Agrobacterium tumefaciens: com  
A:Reference number: S11825; MUID:90301800  
A:Accession: S11839  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1447 <KOB>  
A:Cross-references: EMBL:J03320; NID:9154781; PIDD:AAA91604.1; PTD:9154796  
R:Hagiya, M.; Close, T.J.; Tait, R.C.; Kado, C.I.  
Proc. Natl. Acad. Sci. U.S.A. 82, 2669-2673, 1985  
A:Title: Identification of pTIC58 plasmid-encoded proteins for virulence in Agrobacterium  
A:Reference number: A94037; MUID:85190558  
A:Accession: C22666  
A:Molecule type: DNA  
A:Residues: 167-174; PSMRADONVAPSPGRSLMIPALGSSRLKASVSKTRVNSRRKEITONSANTSH', 'PHLKH  
A:Note: the authors translated the codon ATA for residue 67 as Met and ATA for residue 1  
A:Note: this sequence has been revised in reference S11825  
C:Genetics:  
A:Genome: Plasmid

## Query Match

Best Local Similarity 11.6%; Score 79.5; DB 2; Length 447;  
Best Local Similarity 26.3%; Pred. No. 1.4;  
Matches 35; Conservative 16; Mismatches 43; Indels 39; Gaps 8;

OY 11 VPP-----GRMTQRPGLTYE-----DEGRWTV--VVYRF-----NPSRREMA-- 47  
DB 49 LPDQJHELARSWQETGTDESQPEDEQOELTHIVSPAGTSQVAAYASRWMAAE 108  
OY 48 ---RASGSRPESTIVHL-----WQNAVHTRRELLSSQMPFSQLPAWOLYPGKRYA 98  
DB 109 MFGSGAGGGRYNYLAFHIDRPHLHVVRNRELLGHGWLKIS-----RRHQYNTDA 162  
OY 99 ADSSEWETA-DHG 110  
DB 163 LRKMAEISLRHG 175

## RESULT 4

B37763  
vird2 protein - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens  
C:Date: 31-May-1991 #sequence-revision 31-May-1991 #text-change 08-Oct-1999  
C:Accession: B37763  
R:Wang, K.; Herrera-Estrella, A.; Van Montagu, M.  
J. Bacteriol. 172, 4432-4440, 1990  
A:Title: Overexpression of vird1 and vird2 genes in Agrobacterium tumefaciens enhances

A:Reference number: A37763; MUID:90330550  
A:Accession: B37763

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <MAN>  
A:Cross-references: GB:M33673; NID:9142284; PIDD:AAA2111.1; PTD:9142286

Query Match 11.6%; Score 79.5; DB 2; Length 447;  
Best Local Similarity 26.3%; Pred. No. 1.4;  
Matches 35; Conservative 16; Mismatches 43; Indels 39; Gaps 8;

OY 11 VPP-----GRMTQRPGLTYE-----DEGRWTV--VVYRF-----NPSRREMA-- 47  
DB 49 LPDQJHELARSWQETGTDESQPEDEQOELTHIVSPAGTSQVAAYASRWMAAE 108  
OY 48 ---RASGSRPESTIVHL-----WQNAVHTRRELLSSQMPFSQLPAWOLYPGKRYA 98  
DB 109 MFGSGAGGGRYNYLAFHIDRPHLHVVRNRELLGHGWLKIS-----RRHQYNTDA 162  
OY 99 ADSSEWETA-DHG 110  
DB 163 LRKMAEISLRHG 175

## RESULT 5

S03020  
dopamine beta-monooxygenase (EC 1.14.17.1) precursor - human

N:Alternate names: dopamine beta-hydroxylase  
C:Species: Homo sapiens (man)  
C:Date: 21-Nov-1993 #sequence-revision 10-Nov-1995 #text-change 08-Oct-1999  
C:Accession: S03020; S06283; S61362; I37276  
R:Kobayashi, K.; Kurosawa, Y.; Fujita, K.; Nagatsu, T.  
Nucleic Acids Res. 17, 1089-1102, 1989  
A:Title: Human dopamine beta-hydroxylase gene: two mRNA types having different 3'-ter  
A:Reference number: S03020; MUID:89160241  
A:Accession: S03020  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-603 <KOB>  
A:Cross-references: EMBL:X13255; NID:930473; PIDD:CAA31631.1; PTD:930474  
A:Note: 304-Ser and 535-Cys were also found; the amino acid exchange at residue 535 w  
R:Laumouroux, A.; Vigny, A.; Biguet, N.F.; Darmon, M.C.; Franck, R.; Henry, J.P.; Mall  
EMBO J. 6, 3931-3937, 1987  
A:Title: The primary structure of human dopamine-beta-hydroxylase: insights into the  
A:Reference number: S06283; MUID:89168633  
A:Accession: S06283  
A:Molecule type: mRNA  
A:Residues: 1-196; T, 198-534, C, 536-603 <LAM>  
A:Cross-references: EMBL:Y00096; NID:930455; PIDD:CAA68285.1; PTD:930456  
A:Note: part of this sequence was confirmed by protein sequencing  
A:Note: 502-Leu and 507-Gly were also found  
R:Li, B.; Tsing, S.; Koske, A.H.; Nguyen, B.; Osen, E.G.; Bach, C.; Chan, H.; Barnett  
Biochem. J. 313, 57-64, 1996  
A:Title: Expression of human dopamine beta-hydroxylase in Drosophila Schneider 2 cell  
A:Reference number: S61362; MUID:96132606  
A:Accession: S61362  
A:Molecule type: protein  
A:Residues: 26-34 <LIB>  
C:Superfamily: peptidylglycine monooxygenase I homology  
C:Keywords: catecholamine biosynthesis; copper; glycoprotein; monooxygenase; oxidized  
F:1-25/Domain: signal sequence #status predicted <SID>  
F:26-603/Product: dopamine beta-hydroxylase #status predicted <AMT>  
F:282-505/Domain: peptidylglycine monooxygenase I homology <P6M>

## Query Match

Best Local Similarity 11.6%; Score 79.5; DB 2; Length 603;  
Best Local Similarity 23.8%; Pred. No. 2.1;  
Matches 30; Conservative 19; Mismatches 46; Indels 31; Gaps 5;

OY 3 SEASVRLGPPGLTIQRPGLTYEDEGRWTVVVFNFNSRREWARASOGSRPESTIVH 62  
DB 163 LRKMAEISLRHG 175

Db 377 TDCKTQALDPGSHIFASDLHTLGRKVVTVLR---DGREMLYNODNHSP----- 428

QY 63 LQMAVHTRELSSGQMPFSQLPAVMOLYPCR-----KYPAADSPWEIADHGOIDSE 116

Db 429 -----HFOEI-----RLMKVYVSHPGDVLTSCTVTEDR---ELATVGGFGTILE 471

QY 117 QLVITY 122

Db 472 EMCVNY 477

RESULT 6

C83240

Probable transcription regulator PA3249 [Imported] - Pseudomonas aeruginosa (strain PA01

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83240

R:Stover, C.K.; Pham, X.O.; Eyrvin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, T.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

.. Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: C83240

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-238 <STO>

A:Cross-references: GB:AE004747; GB:AE004091; NID:g9949362; PIDN:AA06637.1; GSPDB:GN001

A:Experimental source: strain PA01

A:Genetics:

A:Gene: PA3249

Query Match 11.5%; Score 79; DB 2; Length 238;

Best Local Similarity 23.9%; Pred. No. 0.75;

Matches 28; Conservative 17; Mismatches 46; Indels 26; Gaps 3;

QY 5 ASVRLGVPPRLMIOIRPGIYEDEGRFTWTVVFRNPSRRMARASGSRYSPTVHL 64

Db 42 ATTRITRLREALIOLESGLYRRE-----RGMFVSPERLAYNPLVRSHF 87

QY 65 QMA-----VHTRELSSGQMPFS-----OLPAVMOLYGRYRAADSPWEIADH 109

Db 88 AMVSEGRVPAETVLSARQMPASAAVCELEPLASVVOIRARVDGRLVLYVEH 144

RESULT 7

E83796

Two-component sensor histidine kinase BH1173 [Imported] - Bacillus halodurans (strain C-

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000

C:Accession: E83796

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20263314

A:Accession: E83796

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <STO>

A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA04892.1; GSPDB:GN00

A:Experimental source: strain C-125

A:Genetics:

A:Gene: BH1173

Query Match 11.0%; Score 75.5; DB 2; Length 476;

Best Local Similarity 27.7%; Pred. No. 4.1;

Matches 36; Conservative 16; Mismatches 53; Indels 25; Gaps 7;

QY 19 QPGRGYE-DEGRFTWTVVFRNPSRR-MAASGSRYSPTVHLQMAVHTRELSS 76

Db 115 QSRGVELSTNGATITVYVTKNSGRGSLISYMDITRDSMVNRLEMLYTL-LISS 173

QY 77 GQMPFSQLPAVMOLY-----LYPGKRAADSF-----WEIAD-----HGOIDSE 117

Db 174 A---LSLTPALMLKHYLRPLILLGNRLQIADNRWKEPPFKWGEDFQKLSNOFEMRQ 230

QY 118 LVITQPERK 127

Db 231 NLVAYDOSQK 240

RESULT 8

S06884

virD protein - Agrobacterium rhizogenes plasmid pRiA4b

C:Species: Agrobacterium rhizogenes

C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 08-Oct-1999

C:Accession: S06884

R:Hirayama, T.; Murakami, T.; Okawa, H.; Oka, A.

Mol. Gen. Genet. 213, 229-237, 1988

A:Title: Organization and characterization of the virD genes from Agrobacterium rhiz

A:Reference number: S06881; MUID:89039712

A:Accession: S06884

A:Molecule type: DNA

A:Residues: 1-436 <HTR>

A:Cross-references: EMBL:X12867; NID:g38995; PIDN:CAA31351.1; PID:g39000

A:Genetics:

A:Gene: virD2

A:Genome: plasmid

Query Match 10.9%; Score 74.5; DB 2; Length 436;

Best Local Similarity 25.6%; Pred. No. 4.7;

Matches 34; Conservative 17; Mismatches 43; Indels 39; Gaps 8;

QY 11 VPP-----GRLMIOIRPGIYE-----DEGRFTWTV--VVYR-----NPSRRMA-- 47

Db 49 LPDDQIHELARSWQETGTVDSDQDERQDELTHIYVFRAGTSQAAMASREMAE 108

QY 48 ---RASGSRYSPTVHL-----WQMAVHTRELSSGQMPFSQLPAVMOLYPGKRYA 98

Db 109 MEGSAGGGSNTYITAFHDHDPHLHVAVNRRLGHGMLKIS-----RRHPOLNYDA 162

QY 99 ADSPWEIA-DHG 110

Db 163 LRLNMAEISLRHG 175

RESULT 9

J00137

hypothetical 30.1K protein - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 07-Jun-1996

C:Accession: J00137

R:Kato, J.; Chu, L.; Kitano, K.; Devault, J.D.; Kimbata, K.; Chakrabarty, A.M.; Misra

Gene 84, 31-38, 1989

A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in

A:Reference number: J00132; MUID:90108714

A:Accession: J00137

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-261 <KAT>

A:Note: 3-Met could also be the initiator

A:Genetics:

A:Start codon: GTG

Query Match 10.7%; Score 73.5; DB 2; Length 261;

Best Local Similarity 27.0%; Pred. No. 3.2;

Matches 30; Conservative 11; Mismatches 41; Indels 29; Gaps 6;

QY 13 PGLMIOIRPGIYE-----DEGRFTWTVVFRNPSRRMARAS-----QGSREP 57

Db 51 PKORRRLPRGMSPORSGQERAMRLROYSLRLVSPQAMPVPSPORSGPAMLAQSR--P 108

QY 58 SITVHLQMAVHTRELLSSGQMFSQLPAWV-----QYEGKRYRADSSP 103  
 Db 109 RVSPHAMPBA-----WLRASRLRFS--PRAPVPYSPASPMALRSLNRF 152

## RESULT 10

probable metallo-oxidoreductase PA3768 [Imported] - Pseudomonas aeruginosa (strain PA01)  
 G83175  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83175  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B.  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lam,  
 .; Loay, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: G83175  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-463 <STO>  
 A:Cross-references: GB:AE004795; GB:AE004091; NID:g9949931; PIDN:AA607155.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA3768

Query Match 10.7%; Score 73; DB 2; Length 463;  
 Best Local Similarity 24.3%; Pred. No. 7.2;  
 Matches 35; Conservative 18; Mismatches 55; Indels 36; Gaps 8;

QY 1 MASEASVRLGVPPGRIMIQRRGI-----YEDEGRITVTVY-----VRF 39  
 Db 12 IAGLAVVGLGAGARLMLARPOVADYELIAPLDLETVPGFSSPALAYGCGPVEL 71  
 QY 40 NPSRRFARASQGSREPSITVHLQMAVHTRELLSSGQMFSQLPAWOLYPR----K 95  
 Db 72 RANQGMELVFRNTNIDPTTH-WH---GTRLPLEMDGVPIYISQPV---QPSGFIYQ 124  
 QY 96 YRAADS-SFWETADHGQIDSEQL 118  
 Db 125 FKTDAGSYWY---HPLMSSEQL 145

## RESULT 11

B25063

hypothetical virD2 protein - Agrobacterium tumefaciens plasmids

C:Species: Agrobacterium tumefaciens  
 C>Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 08-Oct-1999  
 C:Accession: B25063, B25063  
 R:Jayaswal, R.K.; Veluthambi, K.; Gelvin, S.B.; Silghom, J.L.  
 J. Bacteriol. 169, 5035-5045, 1987  
 A:Title: Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molec  
 A:Reference number: A91846; MUID:88032822  
 A:Accession: B25063  
 A:Molecule type: DNA  
 A:Residues: 1-424 <JAY>  
 A:Cross-references: EMBL:M17989; NID:g142288; PIDN:AAA22114.1; PID:g142290  
 A>Note: plasmid pT1A6  
 R:Yanovsky, M.F.; Porter, S.G.; Young, C.; Albright, L.M.; Gordon, M.P.; Nester, E.W.  
 Cell 47, 471-477, 1986  
 A:Title: The virD operon of Agrobacterium tumefaciens encodes a site-specific endonuclease  
 A:Reference number: A90886; MUID:87028239  
 A:Accession: B25063  
 A:Molecule type: DNA  
 A:Residues: 1-424 <YAN>  
 A>Note: plasmid pT1A6NC  
 C:Genetics:  
 A:Gene: virD2  
 A:Genome: plasmid  
 C:Keywords: endonuclease

Query Match 10.6%; Score 72.5; DB 2; Length 424;  
 Best Local Similarity 24.1%; Pred. No. 7.3;  
 Matches 38; Conservative 24; Mismatches 49; Indels 47; Gaps 10;

QY 11 VPPGRILVPPGRIMIQRRGI-----YEDEGRITVTVY-----VRF 39  
 Db 49 VPPDQIRLAQSVTEAGIYDESGSDDDRODITHTIVSPAGTDGTAAYEASREMAE 108  
 QY 50 SGG-----RYEBSITVHL-----WQMAVHTRELLSSGQMFSQLPAWOL-YEGKRYR 97  
 Db 109 MFGSGYGGGRFVNTLTAHVDRDPRHLVHVYVNRRELLGHWKLSIRRP--QLNYGDLKK 166  
 QY 98 ADSSFEIADHGQI-----DSMDLVLTQPERK 127  
 Db 167 MAEISL-----RHGIVDAISRREKRIAEPIYTAERHR 200

## RESULT 12

T34689

probable iron-sulfur binding oxidoreductase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 C:Accession: T34689  
 R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, May 1998  
 A:Reference number: Z21553  
 A:Accession: T34689  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-513 <HAR>  
 A:Cross-references: EMBL:AL023517; PIDN:CA18985.1; GSPDB:GN00070; SCQEDB:SC1B5.11C  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCQEDB:SC1B5.11C

Query Match 10.6%; Score 72.5; DB 2; Length 513;  
 Best Local Similarity 32.8%; Pred. No. 9.2;  
 Matches 19; Conservative 7; Mismatches 19; Indels 13; Gaps 2;

QY 11 VPPGRILVPPGRIMIQRRGI-----YEDEGRITVTVY-----VRF 39  
 Db 436 LPFGEGAVYVRAGGGRILAVYRDEBGLAHVSPRCHLGLCLVFNARAMECPGCSRF 493

## RESULT 13

B70674

probable mbtB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 03-Nov-2000  
 C:Accession: B70674  
 R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Fellous, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: B70674  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1414 <COL>  
 A:Cross-references: GB:Z81371; GB:AL123456; NID:g3261669; PIDN:CAB03756.1; PID:g16573  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: mbtB  
 C:Superfamily: Mycobacterium tuberculosis mbtB protein; acetate--CoA ligase homology;  
 C:Keywords: carrier protein  
 F:5-74/Domain: acyl carrier protein homology <ACP>  
 F:598-1037/Domain: acetate--CoA ligase homology <ACL>  
 F:1060-1131/Domain: acyl carrier protein homology <ACPI>  
 F:1189-1400/Domain: oleoyl-[acyl-carrier protein] hydrolase homology <ACPH>

C;keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Search completed: November 29, 2001, 04:04:10  
Job time: 251 sec

Fri Nov 30 11:03:11 2001

us-09-526-329-39.rpr

---

Page 6



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2001, 04:03:14 ; Search time 9.89 Seconds

(without alignments)  
474,530 Million cell updates/sec

Title: US-09-526-329-39  
Perfect score: 685  
Sequence: 1 MASEASVRLGVPGRLMIOIR.....HGQIDSEQLVLTQPEKRD 128

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	100.0	128	1	TCLB_HUMAN
2	178.5	26.1	116	1	TBL1_MOUSE
3	178.5	26.1	120	1	TBL4_MOUSE
4	178	26.0	107	1	MTC2_MOUSE
5	176.5	25.8	107	1	MTC2_HUMAN
6	175.5	25.6	122	1	TBL3_MOUSE
7	169	24.7	116	1	TCLB_MOUSE
8	163.5	23.9	121	1	TBL5_MOUSE
9	158.5	23.1	117	1	TBL2_MOUSE
10	149	21.8	114	1	TCLB_HUMAN
11	79.5	11.6	447	1	VID2_AGRIS
12	79.5	11.6	603	1	DOPO_HUMAN
13	74.5	10.9	436	1	VID2_AGRIS
14	72.5	10.6	424	1	GLGH_SOLTU
15	70	10.2	366	1	VUL_HSVGU
16	69.5	10.1	775	1	LYS4_EMENT
17	69.5	10.1	1411	1	T297_HUMAN
18	69.5	10.1	1012	1	POLS_IBDVA
19	69	10.1	762	1	PM74_YEAST
20	67.5	9.9	1012	1	POLS_IBDVA
21	67.5	9.9	1012	1	POLS_IBDVA
22	67	9.8	125	1	DHSC_MOUSE
23	67	9.8	3067	1	CAIC_MOUSE
24	66.5	9.7	595	1	VG2B_BPMU5
25	66.5	9.7	1490	1	CRK7_HUMAN
26	66	9.6	326	1	TRM4_BACR
27	66	9.6	535	1	ARKS_PSBAB
28	66	9.6	993	1	POLS_IBDVA
29	66	9.6	1012	1	POLS_IBDVA
30	66	9.6	1012	1	POLS_IBDVA
31	66	9.6	1170	1	DP3A_TREPA
32	65.5	9.6	808	1	SYFA_SYNP7
33	65.5	9.6	877	1	SYA_THIPE

34	65	9.5	286	1	GUB_RHOMR	P45798 rhodothermu
35	65	9.5	1233	1	HCA_OCTDO	P12659 octopus dof
36	64.5	9.4	295	1	TF_RAT	P42533 rattus norv
37	64.5	9.4	439	1	AC48_MOUSE	O97034 mus musculu
38	64.5	9.4	1148	1	MFD_ECOLI	P30958 escherichia
39	64	9.3	310	1	PARB_XYLEA	O96063 xylella fas
40	64	9.3	529	1	LAG3_HUMAN	P18627 homo sapien
41	64	9.3	2410	1	MOR1_SCHPO	O90588 schizosacch
42	63	9.2	380	1	KRIS_HSV11	O00097 ictaluriid h
43	63	9.2	684	1	CDK9_CAEEL	P46551 caenorhadid h
44	63	9.2	1442	1	CPSA_HUMAN	O10570 homo sapien
45	63	9.2	1444	1	CPSA_BOVIN	O10569 bos taurus

## ALIGNMENTS

RESULT 1  
ID TCLB\_HUMAN  
AC O95988;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE T-CELL LEUKEMIA/LYMPHOMA PROTEIN 1B (TCLB ONCOGENE)  
DE (SYNCTIOTROPHOBLAST-SPECIFIC PROTEIN) (STN-1).  
GN TCLB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:99178995; PubMed:1007617;  
RA Pekarsky Y., Hallas C., Isobe M., Russo G., Croce C.M.;  
RT "Abnormalities at 14q32.1 in T cell malignancies involve two  
RT oncogenes";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2949-2951(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Jiang B., Mendelson C.R.;  
RT "A syncytiotrophoblast-specific gene Syn-1 cloned from human  
RT syncytiotrophoblast subcloned cDNA library";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES.  
CC -!- DISEASE: ACTIVATED IN CHRONIC T-CELL LEUKEMIAS (T-CLL) CARRYING  
CC REARRANGEMENTS OF THE 14Q32.1 REGION.  
CC -!- SIMILARITY: BELONGS TO THE TCLB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF110466; AAD16997.1;  
CC EMBL: AF110465; AAD16996.1;  
CC EMBL: AF137027; AAD30130.1;  
CC DR HSSP: P56278; IALX.  
CC DR MIM: 603769;  
CC DR InterPro: IPR002709; TCLB\_MTCPL.  
CC DR ProDom: PD015575; TCLB\_MTCPL.1.  
CC DR Proto-oncogene: Chromosomal translocation.  
CC SEQUENCE 128 AA: 14846 MW: 684233 Da  
CC

Query Match 100.0%; Score 685; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1e-65;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MASEASVRLGVPGRLMIOIRPGIYEDGRTWVTVVVRNPNRREARASOGSRYEEST 60

```

Db 1 MASESVRLVPGRLMIOREITTEDEGGTWTVTYVFNRSREMARASQSGREPSIT 60
    |||
Qy 61 VHLWMAVHTRELLSSGQMPFSQLPVAVMOLYPGKRYRAADSFWEIADHQIDSMDQVL 120
    |||
Db 61 VHLWMAVHTRELLSSGQMPFSQLPVAVMOLYPGKRYRAADSFWEIADHQIDSMDQVL 120
    |||
Qy 121 TYOPEPRD 128
    |||
Db 121 TYOPEPRD 128
    |||

RESULT 2
ID TLBL_MOUSE STANDARD: PRT: 116 AA.
AC P56840:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TCLI1B1 PROTEIN.
GN TCLI1B1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX SEQUENCE FROM N.A.
RP MEDLINE=20056259; PubMed=10588720;
RA Hallas C., Pekarsky Y., Itoyama T., Varnum J., Bichi R.,
RA Rothstein J.L., Croce C.M.;
RT "Genomic analysis of human and mouse TCLI1 loci reveals a complex of
RT tightly clustered genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14418-14423(1999).
CC -1- SIMILARITY: BELONGS TO THE TCLI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF195488; AAF12801.1; -
DR MGD: MGI:1351604; TCLI1B1.
DR InterPro: IPR002709; TCLI_MTCP1.
DR ProDom: PD015575; TCLI_MTCP1; 1.
KW Multigene family.
SQ SEQUENCE 116 AA; 13432 MW; 5D2E08EBC7BE2A64 CRC64;

Query Match 26.1%; Score 178.5; DB 1; Length 116;
Best Local Similarity 34.2%; Pred. No. 3.5e-12;
Matches 41; Conservative 19; Mismatches 51; Indels 9; Gaps 2;

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX SEQUENCE FROM N.A.
RP MEDLINE=20056259; PubMed=10588720;
RA Hallas C., Pekarsky Y., Itoyama T., Varnum J., Bichi R.,
RA Rothstein J.L., Croce C.M.;
RT "Genomic analysis of human and mouse TCLI1 loci reveals a complex of
RT tightly clustered genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14418-14423(1999).
CC -1- SIMILARITY: BELONGS TO THE TCLI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF195492; AAF12805.1; -
DR MGD: MGI:1351604; TCLI1B4.
DR InterPro: IPR002709; TCLI_MTCP1.
DR ProDom: PD015575; TCLI_MTCP1; 1.
KW Multigene family.
SQ SEQUENCE 120 AA; 14123 MW; 7F58E0AEBFF339 CRC64;

Query Match 26.1%; Score 178.5; DB 1; Length 120;
Best Local Similarity 36.4%; Pred. No. 3.7e-12;
Matches 43; Conservative 13; Mismatches 53; Indels 9; Gaps 2;

Qy 6 SVR--LGVPRGLMTIORPGIYEDDEGGTWTVTYVFNRSREMARASQSGREPSITVH 62
    |||
Db 4 SVRFQCFPPFPCVCTRDIDYEDHGRGVAAKVEYS-----SHPYSKETGCVYVH 57
    |||

Qy 63 LWMAVHTRELLSSGQMPFSQLPVAVMOLYPGKRYRAADSFWEIADHQIDSMDQVL 120
    |||
Db 58 LWQMTTLTPEPSDLSKTFNPLRTYRLSRNTYRGADAMRWLRVNSQYGTIELVL 115
    |||

RESULT 4
ID MTC2_MOUSE STANDARD: PRT: 107 AA.
AC Q60945:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE P13 MTC2P-1 PROTEIN (MATURE T-CELL PROLIFERATION-1 TYPE B1) (MTC2P-1
DE TYPE B1) (P13MTC2P1).
GN MTC2P1 OR C6.1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX SEQUENCE FROM N.A.
RP TISSUE=T-cell;
RC MEDLINE=96202516; PubMed=8634440;
RA Madani A., Choukroun V., Soulier J., Cacheux V., Claissé J.-F.,
RA Valensi F., Daliphard S., Cazin B., Lévy V., Leblond V.,
RA Daniel M.-T., Sigaux F., Stern M.-H.;
RT "Expression of p13MTC2P1 is restricted to mature T-cell proliferations
RT with t(4;14) translocations.";
RL Blood 87:1923-1927(1996).
CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING PRODUCES TWO SHORT, YET
CC COMPLETELY DIFFERENT PROTEINS. THE LONGER PRODUCT, TYPE-B1, IS
CC KNOWN AS P13 MTC2P-1 (AC Q60945). THE SHORTER PRODUCT, TYPE-A, IS
CC -1- TISSUE SPECIFICITY: NOT FOUND AT A SIGNIFICANT LEVEL IN ANY
CC TISSUE.

```



KW Multigene family.  
SQ SEQUENCE 122 AA: 141170 MW: BD1501F81C24F230 CRC64:

Query Match 25.6%; Score 175.5; DB 1; Length 122;  
Best Local Similarity 36.7%; Pred. No. 7,8e-12;  
Matches 40; Conservative 12; Mismatches 50; Indels 7; Gaps 1;

OY 12 PPRFVLCRBDIDIBENGKQWYAKV-----ETSSPRGSRIFETCTVLMQHTTQ 71  
DB 13 PPRFVLCRBDIDIBENGKQWYAKV-----ETSSPRGSRIFETCTVLMQHTTQ 65  
OY 72 ELLSSQMFPSQLPAWOLYPRGKRYRADSSFWETADHGOIDSMQVLY 120  
DB 66 EPPPOPTINNLSLPTWRLWRLESNMTYTGDTGYWRLDLDSQMGDTQLLL 114

RESULT 7

TCL1\_MOUSE STANDARD: PRT: 116 AA.

AC P56280;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE T-CELL LEUKEMIA/LYMPHOMA PROTEIN 1A (P14 TCL1 PROTEIN) (TCL1 ONCOGENE)  
DE (TCL-1 PROTEIN).  
GN TCL1 OR TCL1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97430049; PubMed=9285687;  
RA Narducci M.G., Virgilio L., Engles J.B., Buchberg A.M., Billips L.,  
RA Facchiano A., Croce C.M., Russo G., Rothstein J.L.;  
RT "The murine Tcl1 oncogene: embryonic and lymphoid cell expression.";  
RL Oncogene 15:919-926(1997).  
CC -1- SIMILARITY: BELONGS TO THE TCL1 FAMILY.  
CC -1- SUBCELLULAR LOCATION: MICROSOMAL FRACTION (BY SIMILARITY).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF031956; AAB87461.1; -.  
DR EMBL: Y15376; CAAY559.1; -.  
DR HSSP: P56279; USG.  
DR MGD: MGI:1097166; Tc11.  
DR InterPro: IPR002709; TCL1\_MTCPL.  
DR Pfam: PF01840; TCL1\_MTCPL; 2.  
DR ProDom: PD015575; TCL1\_MTCPL; 1.  
KW Microsome.  
SQ SEQUENCE 116 AA: 14112 MW: 40DBED2F973F389A CRC64:

Query Match 24.7%; Score 169; DB 1; Length 116;  
Best Local Similarity 32.8%; Pred. No. 3,6e-11;  
Matches 40; Conservative 20; Mismatches 44; Indels 18; Gaps 2;

OY 1 MASASVRLGVP--DGRLLIORPGIYEDERGRTWTVVFRNPSRREMARASQGSRYEPS 58  
DB 1 MAIRRAHRAETFAHNRRLMTWEKIVYLDERFSWLPVYIKN-----EK 44  
OY 59 ITVHLWQMAVHRELLSSQMFPSQLPAWOLYPRGKRYRADSSFWETADHGOIDSMQVLY 118  
DB 45 FOVTLROEDVTGLGAMSPSQLPVPELPLMWOLYPRDVRSCDSWYQVLIYHFKFVEDM 104  
OY 119 VL 120

DB 105 LL 106

RESULT 8  
TCL1\_MOUSE STANDARD: PRT: 121 AA.

AC P56845;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TCL1B5. PROTEIN.  
GN TCL1B5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20056259; PubMed=10588720;  
RA Hallas C., Pekarsky Y., Itoyama T., Varnum J., Bichi R.,  
RA Rothstein J.L., Croce C.M.;  
RT "Genomic analysis of human and mouse TCL1 loci reveals a complex of  
RT tightly clustered genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:14418-14423(1999).  
CC -1- SIMILARITY: BELONGS TO THE TCL1 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF195493; AAF12806.1; -.  
DR MGD: MGI:1351635; Tc1b5.  
DR InterPro: IPR002709; TCL1\_MTCPL.  
DR ProDom: PD015575; TCL1\_MTCPL; 1.  
KW Multigene family.  
SQ SEQUENCE 121 AA: 13668 MW: 32815CDD629B4297 CRC64:

Query Match 23.9%; Score 163.5; DB 1; Length 121;  
Best Local Similarity 33.0%; Pred. No. 1,4e-10;  
Matches 37; Conservative 14; Mismatches 44; Indels 17; Gaps 2;

OY 19 QRP-----GTYDEDEGRRTWTVVFRNPSRREMARASQGSRYEPSITVHLWQNAV 68  
DB 9 QRLPLVLYSVSLGIYEDENHRVIAVNV-----ETSSHSGNRIFETCTVLMQHTTQ 61  
OY 69 HTEELLSSQMFPSQLPAWOLYPRGKRYRADSSFWETADHGOIDSMQVLY 120  
DB 62 LPDEPPPOPTINNLSLPTWRLWRLESNMTYTGDTGYWRLDLDSQMGDTQLLL 113

RESULT 9

TCL1\_MOUSE STANDARD: PRT: 117 AA.

AC P56841;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TCL1B2. PROTEIN.  
GN TCL1B2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20056259; PubMed=10588720;  
RA Hallas C., Pekarsky Y., Itoyama T., Varnum J., Bichi R.,

RA Rothstein J.L., Croce C.M.;  
 RT "Genomic analysis of human and mouse TCL1 loci reveals a complex of  
 RT tightly clustered genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14418-14423(1999).  
 CC -1- SIMILARITY: BELONGS TO THE TCL1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF195489; AAF12802.1; -  
 DR MGD: MGI:1351609; Tc1b2.  
 DR InterPro: IPR002709; TCL1\_MTCPL.  
 DR ProDom: PD015575; TCL1\_MTCPL; 1.  
 KW Multigene family;  
 SO SQUENCE 117 AA; 13672 MW; A6B0851165E9B0AA CRC64;

Query Match 23.1%; Score 158.5; DB 1; Length 117;  
 Best Local Similarity 31.0%; Pred. No. 4.7e-10;  
 Matches 36; Conservative 17; Mismatches 52; Indels 11; Gaps 2;

OY 10 GYPRGLMIQ-----RGIYEDEGRWTWYVVRNPSRREMARASQGSRYPSITVHLN 64  
 DB 5 GYPRRLPOLVLISTGPGFEDHHRIMWAKL-----ETCSHSPYCNKLEICVTHLM 58  
 OY 65 QNAVHRELSSGMPESQLPVAVOLYPRKRYRADSFMEIADHQIDSMQVLYTOE 120  
 DB 59 QMTRYPDEPAPYPMVNYNPLPMTWRLASMTYRGTDAMGMRLNHSQGYDTYDLIL 114

RESULT 10  
 ID TCL1\_HUMAN STANDARD; PRT; 114 AA.  
 AC P56279;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE T-CELL LEUKEMIA/LYMPHOMA PROTEIN 1A (P14 TCL1 PROTEIN) (TCL1 ONCOGENE)  
 DE (TCL-1 PROTEIN).  
 GN TCL1A OR TCL1.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95107991; PubMed=7809072;  
 RA Virgilio L., Narducci M.G., Isobe M., Billips L.G., Cooper M.D.,  
 RA Croce C.M., Russo G.;  
 RT "Identification of the TCL1 gene involved in T-cell malignancies";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12530-12534(1994).  
 RN [2]  
 RP CHARACTERIZATION AND LOCALIZATION.  
 RX MEDLINE=95079394; PubMed=7987816;  
 RA Fu T.-B., Virgilio L., Narducci M.G., Facchiano A., Russo G.,  
 RA Croce C.M.;  
 RT "Characterization and localization of the TCL-1 oncogene product";  
 RL Cancer Res. 54:6297-6301(1994).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=98179932; PubMed=9519406;  
 RA Hob F., Yang Y.-S., Guignard L., Padilla A., Stern M.-H.,  
 RA Ihosie I.-M., van Tilbourn H.;  
 RT "Crystal structure of p14TCL1, an oncogene product involved in T-cell  
 RT prolymphocytic leukemia, reveals a novel beta-barrel topology";  
 RL Structure 6:147-155(1998).  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL FRACTION.

CC -1- TISSUE SPECIFICITY: RESTRICTED IN THE T-CELL LINEAGE TO IMMATURE  
 CC THYMOCYTES AND ACTIVATED PERIPHERAL LYMPHOCYTES. PREFERENTIALLY  
 CC EXPRESSED EARLY IN T- AND B-LYMPHOCYTE DIFFERENTIATION.  
 CC -1- DISEASE: ACTIVATED IN CHRONIC T-CELL LEUKEMIAS (T-CLL) CARRYING  
 CC T(14;14)(Q11;Q32) OR T(7;14)(Q35;Q32) CHROMOSOME TRANSLOCATIONS OR  
 CC A INV(14)(Q11;Q32) CHROMOSOME INVERSION THAT INVOLVES THE T-CELL  
 CC RECEPTOR ALPHA/DELTA LOCUSSES.  
 CC -1- SIMILARITY: BELONGS TO THE TCL1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X82240; CAA57708.1; -  
 DR PDB: 1JSG; 1B-MAR-98.  
 DR MIM: 186960; -  
 DR InterPro: IPR002709; TCL1\_MTCPL.  
 DR Pfam: PF01840; TCL1\_MTCPL; 2.  
 DR ProDom: PD015575; TCL1\_MTCPL; 1.  
 KW Proto-oncogene; Chromosomal translocation; Microsome; 3D-structure.  
 SO SQUENCE 114 AA; 13459 MW; 90D55ABC97C36D04 CRC64;

Query Match 21.8%; Score 149; DB 1; Length 114;  
 Best Local Similarity 29.2%; Pred. No. 4.6e-09;  
 Matches 33; Conservative 21; Mismatches 45; Indels 14; Gaps 1;

OY 13 PERLTIORPGIYDEGRWTWYVVRNPSRREMARASQGSRYPSITVHLNOMAVHRE 72  
 DB 15 PDRIMAEKEVYLDDEKOHAMLPTEIKRLQ-----LRVLLREDVYAGR 60  
 OY 73 LLSGMPESQLPVAVOLYPRKRYRADSFMEIADHQIDSMQVLYTOE 125  
 DB 61 PMTPYQIGPSLPLIMMQLYDGRYRSSDFMNLVYHKLIDGVEDMLLELDD 113

RESULT 11  
 ID V1D2\_AGR75 STANDARD; PRT; 447 AA.  
 AC P18592; P06521; P06522;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE T-DNA BORDER ENDONUCLEASE V1D2 (Ec 3.1.-.-).  
 GN V1D2.  
 OS Agrobacterium tumefaciens.  
 OS Plasmid pTiC58.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90301800; PubMed=2194232;  
 RA Rosowsky P.M., Powell B.S., Shirasu K., Lin T.-S., Morel P.,  
 RA Zyprian E.M., Steck T.R., Kado C.I.;  
 RT "Molecular characterization of the vir regulon of Agrobacterium  
 RT tumefaciens: complete nucleotide sequence and gene organization of  
 RT the 28.65-kbp regulon cloned as a single unit";  
 RL Plasmid 23:85-106(1990).  
 RN [2]  
 RP PRELIMINARY SEQUENCE FROM N.A.  
 RX MEDLINE=85190558; PubMed=2986128;  
 RA Hagiya M., Close T.J., Tait R.C., Kado C.I.;  
 RT "Identification of pTiC58 plasmid-encoded proteins for virulence in  
 RT Agrobacterium tumefaciens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2669-2673(1985).  
 CC -1- FUNCTION: TUMOR FORMATION BY A TUMEFACIENS INVOLVES THE TRANSFER  
 CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA  
 CC INTO THE PLANT NUCLEAR GENOME. THE VIRD OPERON ENCODES A SITE-



```

RT      "The vird operon of Agrobacterium tumefaciens encodes a site-specific
RT      endonuclease."
RL      Cell 47:471-477(1986).
RN      [2]
RP      SOURCE FROM N.A.
RA      MEDLINE=8032822; PubMed-2922660;
RA      Jayaswal R.K., Veluthambi K., Gelvin S.B., Slightom J.L.;
RT      "Double-stranded cleavage of T-DNA and generation of single-stranded
RT      border-specific endonuclease of Agrobacterium tumefaciens."
RL      J. Bacteriol. 169:5035-5045(1987)
CC      -1- FUNCTION: TUMOR FORMATION BY A TUMORAGENIC INVOLVES THE TRANSFER
CC      AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC      INTO THE PLANT NUCLEAR GENOME. THE VIRD OPERON ENCODES A SITE-
CC      SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC      24 BP DIRECT REPEATS FLANKING THE T-DNA.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed, usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      -----
DR      EMBL AF242881; AAA98390.1; -.
DR      EMBL M17989; AAA22114.1; -.
DR      PIR: B29826; B29826.
DR      PIR: B25063; B25063.
KW      Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;
KW      T-DNA.
SQ      SEQUENCE 424 AA; 47546 MW; B2CB7B82C01A6C1C CRC64;

Query Match 10.68; Score 72.5; DB 1; Length 424;
Best Local Similarity 24.18; Pred. No. 2.7;
Matches 38; Conservative 24; Mismatches 49; Indels 47; Gaps 10;

Oy      11 VPPGRL-----WIOPGTVE-----DESGRTWV--VVYF-----NPSREKRA 49
Db      49 VPPQDRELQSWTENGIVDESGDDDDQDDITHIYSPAGDGTATAYENSKEMAAE 108
Oy      50 SGG-----HYESTIVHL-----WQAVHRELLSSQMPFSQDLAVWQI-YGRKYR 97
Db      109 MFGSGYGGGKNTLTAHVHVDRLPHLVVYVRRRLGHGMLKISRHP--QLNYGLRKK 166
Oy      98 AADSGEWELADHGQI-----DSMEQIVLYTTPERK 127
Db      167 MAEISL-----RGQIVLDTSRAGRGIAERPIITYAEHRR 200

RESULT 15
GLGH.SOLTU STANDARD; PRT; 861 AA.
CD      P30924.
DT      01-JUL-1993 (rel. 26, Created)
DT      01-NOV-1995 (rel. 32, Last sequence update)
DT      20-AUG-2001 (rel. 40, Last annotation update)
DE      1,4-ALPHA-GLUCAN BRANCHING ENZYME (BC 2.4.1.18) (STARCH BRANCHING
DE      ENZYME) (O-ENZYME).
GN      SBE1 OR SBE.
OS      Solanum tuberosum (Potato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; eustersierii I; Solanales; Solanaceae; Solanum.
OC      NCBI_TaxID=4113;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN-CV, DIANELLA;
RA      MEDLINE=94105324; PubMed-8278528;
RA      Poulsen P., Kreiberg J.D.;
RT      "Starch branching enzyme cDNA from Solanum tuberosum."
RT      Plant Physiol. 102:1053-1054(1993)

```



RN [2]  
 RP SEQUENCE OF 279-527 FROM N.A.  
 RC STRAIN-CV, DESIREE; TISSUE-Tuber;  
 RX MEDLINE=92079917; PubMed=1745241;  
 RA Kossmann J., Vissler R.G.F., Mueller-Roeber B., Willmitzer L.,  
 RA Sonnwald U.;  
 RT Cloning and expression analysis of a potato cDNA that encodes  
 RT branching enzyme: evidence for co-expression of starch biosynthetic  
 RT genes.";  
 RL Mol. Gen. Genet. 230:39-44(1991).  
 CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF  
 CC STARCH.  
 CC -1- PATHWAY: THIRD STEP IN STARCH BIOSYNTHESIS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: AMYLOPLAST.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X69805; CAA49463.1; -  
 DR PIR: S18594; S18594.  
 DR Mendel: 13376; SOLtu:Sbel;1.  
 DR InterPro: IPR000461; Alpha\_amylase.  
 DR Pfam: PF00128; alpha-amylase; 1.  
 KW Starch biosynthesis; Transferase; Glycosyltransferase; Amyloplast.  
 FT ACT\_SITE 424 424 BY SIMILARITY.  
 FT ACT\_SITE 484 484 BY SIMILARITY.  
 FT ACT\_SITE 553 553 BY SIMILARITY.  
 SO SEQUENCE 861 AA; 99083 MW; F3D519AC7CF1BEF2 CRC64;

Query Match 10.2%; Score 70; DB 1; Length 861;  
 Best Local Similarity 24.2%; Pred. No. 11;  
 Matches 31; Conservative 17; Mismatches 34; Indels 46; Gaps 8;  
 QY 17 WIORPGIYDEGCRWTVVVRFPNSRRRARA-SOGSRYEPSTIVHLQMAVHTRE-- 72  
 DB 620 WIDFP-----REGNNW-----SYDKCRQWNLADSEHLRYK--FWNAFDRANSLDEKF 666  
 QY 73 -LSSGQMPFS-----QLPAYWQIVGR-----KYRAA-DSSF 103  
 DB 667 SFLASGKQIVSSMDNDKNVYFERGDLVFNFFHKNTYEGYKVGCDLPKRYVALDSDA 726  
 QY 104 WEIADHGQ 111  
 DB 727 WEFQGHGR 734

Search completed: November 29, 2001, 04:06:42  
 Job time: 208 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2001, 04:03:39 ; Search time 23.95 Seconds

(Without alignments)  
781.748 Million cell updates/sec

Title: US-09-526-329-39

Perfect score: 685

Sequence: 1 MASEASVRLGVPGRGLMIQR.....HGQIDSMQVLVTPPERKD 128

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_17:\*  
1: sp\_Archaea:\*  
2: sp\_Bacteria:\*  
3: sp\_Fungi:\*  
4: sp\_Human:\*  
5: sp\_Invertebrate:\*  
6: sp\_Mammal:\*  
7: sp\_mhc:\*  
8: sp\_Organella:\*  
9: sp\_Phage:\*  
10: sp\_Plant:\*  
11: sp\_Rodent:\*  
12: sp\_Virus:\*  
13: sp\_Vertebrate:\*  
14: sp\_Unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score by the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	677	98.8	128	4 Q9UBQ4	Q9UBQ4 homo sapien
2	113	16.5	107	11 Q9QXN9	Q9QXN9 mus musculu
3	81	11.8	448	2 Q9P587	Q9P587 agrobacteri
4	79.5	11.6	447	2 Q44461	Q44461 agrobacteri
5	79.5	11.6	447	2 Q9R6B9	Q9R6B9 agrobacteri
6	79	11.5	238	2 Q9H721	Q9H721 pseudomonas
7	76.5	11.2	588	12 Q70681	Q70681 sugarcane s
8	75.5	11.1	3742	8 Q9TKW8	Q9TKW8 nephroselin
9	75.5	11.1	476	2 Q9KDN8	Q9KDN8 bacillus ba
10	75	10.7	986	4 Q9ULF0	Q9ULF0 homo sapien
11	73	10.7	463	2 Q9HXM7	Q9HXM7 pseudomonas
12	73	10.7	464	10 Q9F199	Q9F199 aradidopsis
13	73	10.7	513	2 Q9L803	Q9L803 pseudomonas
14	73	10.7	2229	2 Q9EMP5	Q9EMP5 streptomyce
15	72.5	10.6	513	2 Q69839	Q69839 streptomyce
16	72.5	10.6	1414	2 P71717	P71717 mycobacteri
17	70	10.2	367	2 Q9KAV7	Q9KAV7 vibrio chol
18	70	10.2	477	12 Q9ILY2	Q9ILY2 beet mosaic
19	70	10.2	830	10 Q04864	Q04864 solanum tub

20	70	10.2	1012	12 Q9WR38	Q9WR38 infectious
21	69.5	10.1	363	12 Q9CJ57	Q9CJ57 human herpe
22	69.5	10.1	433	12 Q9WT62	Q9WT62 human herpe
23	69.5	10.1	477	2 Q9FAX3	Q9FAX3 flexibacter
24	69.5	10.1	1927	5 Q25142	Q25142 halocynthia
25	69	10.1	363	2 Q9Z514	Q9Z514 streptomyce
26	69	10.1	1012	12 Q82628	Q82628 streptomyce
27	69	10.1	1012	12 Q9QPP1	Q9QPP1 infectious
28	69	10.1	1553	5 Q9N859	Q9N859 leishmania
29	69	10.1	1747	5 Q25265	Q25265 lymphatria d
30	69	10.1	1747	5 Q25265	Q25265 lymphatria d
31	68.5	10.0	620	5 Q9N9D4	Q9N9D4 megaselia a
32	68.5	10.0	785	2 Q9K3N5	Q9K3N5 streptomyce
33	68	9.9	567	3 Q9WUS8	Q9WUS8 drosophila
34	68	9.9	567	3 Q9WUS8	Q9WUS8 drosophila
35	68	9.9	747	2 Q99244	Q99244 streptococ
36	68	9.9	1012	12 Q82635	Q82635 infectious
37	68	9.9	1012	12 Q98099	Q98099 infectious
38	68	9.9	1012	12 Q98101	Q98101 infectious
39	68	9.9	1012	12 Q9KCF1	Q9KCF1 infectious
40	68	9.9	1012	12 Q9KCF1	Q9KCF1 infectious
41	68	9.9	1013	12 Q9KCF1	Q9KCF1 infectious
42	68	9.9	1409	3 Q74298	Q74298 penicillium
43	67	9.8	243	10 Q9CAX0	Q9CAX0 arabidopsis
44	67	9.8	390	4 Q9BVF0	Q9BVF0 homo sapien
45	67	9.8	412	4 Q9NXP4	Q9NXP4 homo sapien

## ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	128 AA.
ID Q9UBQ4			
AC Q9UBQ4			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE TGL1 / MRCPL-LIKE 1.			
GN TGL1.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=9274532; PubMed=10344735;			
RA Sugimoto J., Hatakeyama T., Narducci M.G., Russo G., Isobe M.;			
RT Identification of the TGL1/MRCPL-Like 1 (TGL1) gene from the region			
RL next to the TGL1 locus.;			
DR Cancer Res. 59:2313-2317(1999).			
DR EMBL; AB025274; BAA82476.1; -			
DR EMBL; AB025272; BAA82476.1; JOINED.			
DR EMBL; AB018563; BAA76712.1; -			
DR HSSP; P56278; IAI1X.			
DR InterPro; IPR002709; TGL1_MRCPL.			
DR ProDom; PD015575; TGL1_MRCPL.1.			
SO SEQUENCE 128 AA; 14545 MW; 6BA2CF22F62CFB4D CRC64;			

Query Match	98.8%;	Score 677;	DB 4;	Length 128;
Best Local Similarity	99.2%;	Pred. No. 1.3e-62;		
Matches 127;	Conservative	0;	Mismatches	1;
			Indels	0;
			Gaps	0;
QY 1 MASEASVRLGVPGRGLMIQRPGIYEDDEGRMTVYVRFNPSRREKARASGCSRYEPT 60				
DB 1 MASEASVRLGVPGRGLMIQRPGIYEDDEGRMTVYVRFNPSRREKARASGCSRYEPT 60				
QY 61 VHLKQAVATRRLLSSGQMPFSQLPAVWOLYPRKRYRAADSFWELADHGOIDSMQVLV 120				
DB 61 VHLKQAVATRRLLSSGQMPFSQLPAVWOLYPRKRYRAADSFWELADHGOIDSMQVLV 120				



DB 163 LRKMAEISLRHG 175

RESULT 5

ID 09R6B9 PRELIMINARY; PRT; 447 AA.

AC 09R6B9

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE TIORF156 PROTEIN.

GN TIORF156

OS Agrobacterium radiobacter.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OX NCBI\_TaxID=358;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RX MEDLINE=20184752; PubMed=10721727;

RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,

RA Katoh A., Yoshida K.;

RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";

RL Gene 242:331-336(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RX MEDLINE=98193120; PubMed=9524202;

RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;

RT "Novel structural difference between nopaline- and octopine- type trbJ

RT gene: construction of genetic and physical map and sequencing of

RT trbJ/trai and rep genes clusters of a new Ti plasmid pTi-SAKURA.";

RL Blochm. Biophys. Acta 1396:1-7(1998).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;

RT "Genome structure of pTi-SAKURA(1): Strategy for DNA sequencing of a

RT Japanese cherry-Ti plasmid.";

RL Nucleic Acids Symp. Ser. 37:159-160(1998).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;

RT "Genome structure of pTi-SAKURA (111): Characteristics of T-DNA.";

RL Nucleic Acids Symp. Ser. 39:185-186(1998).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;

RT "Genome structure of pTi-SAKURA (1V): Characteristics of tra region.";

RL Nucleic Acids Symp. Ser. 39:187-188(1998).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN-MAF301001;

RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;

RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of

RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.";

RL Nucleic Acids Symp. Ser. 39:265-266(1998).

DR EMBL: AB016260; DAA87781.1; -.

KW plasmid.

SO SEQUENCE 447 AA; 49603 MW; 5A9E646F2D8894AC CRC64;

Query Match 11.6%; Score 79.5; DB 2; Length 447;

Best Local Similarity 26.3%; Pred. No. 3.4;

Matches 35; Conservative 16; Mismatches 43; Indels 39; Gaps 8;

DB 49 LPDQJHEIARSWOETGYDESQDDEKQOELTTHITIVSFAGTSQVAAAYASREMAAE 108

OY 48 ---RASQGSRYEPSTVHL-----WQAVHRELLSSGQMPESQLPVAMOLYPRKRYA 98  
 DB 109 MFGSAGGGRNYLAFHIDRPHLHVNNRELLGHWLKIS-----RRHPOLANDA 162  
 OY 99 ADSSFEIA-DHG 110  
 DB 163 LRKMAEISLRHG 175

RESULT 6

ID 09HYZ1 PRELIMINARY; PRT; 238 AA.

AC 09HYZ1

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE PROBABLE TRANSCRIPTIONAL REGULATOR.

GN PA3249.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Gollery L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Gollery L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Gollery L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

RN [1]

Query Match 11.5%; Score 79; DB 2; Length 238;

Best Local Similarity 23.9%; Pred. No. 1.8;

Matches 28; Conservative 17; Mismatches 46; Indels 26; Gaps 3;

DB 5 ASVRLGVPGRMIORGIYDEGRTWTVVFNPSRRERARASQGSRYEPSTVHLM 64

DB 42 ATTRITLREALIOLISOGILYREE-----RRGWEVSPERLAVNPLVRSHEH 87

OY 65 OMA-----VHRELLSSGOMPFS-----OLPVMOLYGRKRYRAAASWETADH 109

DB 88 AMVSEQGRVATEVLSARQMPASANCELLPULSVYQIRARVQGRILVYVEH 144

RESULT 7

ID 070681 PRELIMINARY; PRT; 588 AA.

AC 070681

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE COAT PROTEIN (FRAGMENT).

GN Sugarcane streak mosaic virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;

OC Polyvirus.

OX NCBI\_TaxID=53954;

RN [1]

RP SEQUENCE FROM N.A.  
 RA Hall J.S., Adams B., Parsons T.J., French R., Lane L., Jensen S.G.;  
 RL Mol. Phylogenet. Evol. 0:0-0(1998).  
 DR EMBL: U75456; AAC16771.1; -  
 DR InterPro: IPR001205; RNA\_POL\_P3D.  
 DR InterPro: IPR001592; Poly\_coat.  
 DR Pfam: PF00680; RNA\_dep-RNA\_pol.1.  
 DR Pfam: PF00767; Poly\_coat.1.  
 KW Coat protein.  
 FT NON TER  
 SQ SEQUENCE 588 AA; 66335 MW; 3876EF59189EA32 CRC64;

Query Match 11.2%; Score 76.5; DB 12; Length 588;  
 Best Local Similarity 32.8%; Pred. No. 9.5;  
 Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 1;

OY 24 YEDEGRMTVTVVVFNPNSRREWARA-----SOGSRYSPTVHLMQMAVTRRE 72  
 DB 11 YERQAGSHVTVGINKFNCNGDELAARFPHNKFKIDADGSRDSSLPLPFLFNCVLIHRE 68

RESULT 8  
 O9TKW8 PRELIMINARY; PRT; 3742 AA.  
 AC O9TKW8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL CHLOROPLAST RF2.  
 GN YCF2.  
 OS Nephroselmis olivacea.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
 OC Chlorodendales; Chlorodendraceae; Nephroselmis.  
 OX NCBI\_TaxID=31312;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=9398694; PubMed=10468594;  
 RA Tunnel M., Otis C., Lemieux C.;  
 RT "The complete chloroplast DNA sequence of the green alga Nephroselmis  
 olivacea: Insights into the architecture of ancestral chloroplast  
 genomes.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 96:10248-10253(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tunnel M., Otis C., Lemieux C.;  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF137379; AAD54848.1; -  
 DR InterPro: IPR003593; AAA.  
 DR InterPro: IPR003599; AAA-subfam.  
 DR Pfam: PF00004; AAA.1.  
 DR SMART: SM00382; AAA.1.  
 KW Chloroplast.  
 SQ SEQUENCE 3742 AA; 436432 MW; 1F45143EDBC89275 CRC64;

Query Match 11.1%; Score 76; DB 8; Length 3742;  
 Best Local Similarity 29.1%; Pred. No. 92;  
 Matches 37; Conservative 12; Mismatches 46; Indels 32; Gaps 8;

OY 14 GRLLIOPR---GIYDEGRMTVTVVVFNPNSRREWARA-----SOGSRYSPTVHLMQMAVTR 70  
 DB 2934 GRLLDDYPHSHSVKSTIEGVDVTVTVL---SVDRQWVHSTLKTIDPFV-----NRNAXYA 2987  
 OY 71 ---RELLSS-----GQMPFSQLPVWQLYPGRK---YRAADSEFWELADHGQIDSM 115  
 DB 2988 RAGRALSSIVSGYGVGILFYSORPR-----GRKAPETYSKIDQKAVE--DHFNAEEC 3039  
 OY 116 EQVLVRY 122  
 DB 3040 KOSTILH 3046

RESULT 9  
 O9KDN8 PRELIMINARY; PRT; 476 AA.  
 AC O9KDN8;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE TWO-COMPONENT SENSOR HISTIDINE KINASE.  
 GN BH1173.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=8665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSOR TRANSDUCTION HISTIDINE  
 KINASES.

CC EMBL: AP001511; BAB04892.1; -  
 DR InterPro: IPR000410; BcTrl\_sensur.  
 DR InterPro: IPR000658; DUF5.  
 DR InterPro: IPR003660; HAMP.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR003661; His\_kina.  
 DR Pfam: PF0672; DUF5.1.  
 DR Pfam: PF02518; HATPase\_c.1.  
 DR Pfam: PF00512; signal.1.  
 DR PRINTS: PR00344; BCTRLSENSOR.  
 DR SMART: SM00304; HAMP.1.  
 DR SMART: SM00387; HATPase\_c.1.  
 DR SMART: SM00388; HisKA.1.  
 KW Complete proteome; kinase; phosphorylation; sensory transduction;  
 KW Transference.  
 SQ SEQUENCE 476 AA; 55661 MW; 0621E01733E0A319 CRC64;

Query Match 11.0%; Score 75.5; DB 2; Length 476;  
 Best Local Similarity 27.7%; Pred. No. 9.4;  
 Matches 36; Conservative 16; Mismatches 53; Indels 25; Gaps 7;

OY 19 QRGQITE-DEGRMTVTVVVFNPNSRRE-WARASGSRYSPTVHLMQMAVTRRELLSS 76  
 DB 115 QSRGRELSTYNGATIFYVTRKNSFGESGLYSTIMMDTYRDSVNRMLRLLYTL-LSS 173  
 OY 77 GQMPFSQLPVWQ-----LYPRKRYRADSSF-----WEIAD-----HGQIDSMQ 117  
 DB 174 A--LSPALPAWLKHYLRQPLILLGNRLRLOIADNRNKKPEPKWGEDPQKLSNFERMRQ 230  
 OY 118 LVLYQPERK 127  
 DB 231 NLVRYDOSK 240

RESULT 10  
 O9ULFO PRELIMINARY; PRT; 986 AA.  
 AC O9ULFO;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE KIAA1270 PROTEIN (FRAGMENT).  
 GN KIAA1270.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=20039619; PubMed=10574462;  
 RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XV.  
 The complete sequences of 100 new cDNA clones from brain which code  
 for large proteins in vitro."  
 RL DNA Res. 6:337-343(1999).  
 DR EMBL: AB033096; BAA86584.1;  
 DR InterPro: IPR002106; AA\_tRNA\_Ligase\_II.  
 DR InterPro: IPR002318; tRNA-synt\_2c.  
 DR Pfam: PF01411; tRNA-synt\_2c; 1.  
 DR PRINTS: PR00980; TRNASYNTHALA.  
 DR PROSITE: PS00339; AA\_tRNA\_Ligase\_II\_2; UNKNOWN\_1.  
 FT NON\_TER 1  
 SQ SEQUENCE 986 AA; 107427 MW; E6FDD207907AA831 CRC64;

Query Match 10.9%; Score 75; DB 4; Length 986;  
 Best Local Similarity 26.0%; Pred. No. 25;  
 Matches 34; Conservative 16; Mismatches 51; Indels 30; Gaps 8;

OY 10 GVPGRMTD-----RGTEDEDEGR-TWYV-----VYREMPSRREKARASGSKRYE 56  
 DB 162 GIPERLMTISYEDGPKAGLPDLETDITLWISGVASRVLSFGPOENFWEKGTDP-CG 220  
 OY 57 PSIVHLMQAVHTELLSSGMPFSQLPAYWQLYPKRYRAADSFWEL----ADHGOI 112  
 DB 221 PCTEIH-YDLA-----GGVGPQLVELMNLVFMQHNENAGSLQLPQRHVDIGM- 269  
 OY 113 DSMQQLVLYQ 123  
 DB 270 -GLERLVAVLY 279

RESULT 11  
 O9HXMT PRELIMINARY; PRT; 463 AA.  
 AC O9HXMT;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PROBABLE METALLO-OXIDOREDUCTASE.  
 GN PA3768.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Gabber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Reiter J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004795; AAG07155.1;  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR002355; MulticCu\_oxidase2.  
 DR Pfam: PR00394; Cu-oxidase; 2.  
 DR PROSITE: PS00079; MULTICOPPER\_OXIDASE; 1.  
 DR PROSITE: PS00080; MULTICOPPER\_OXIDASE2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 463 AA; 51448 MW; 44222B5DFF9EB33 CRC64;

Query Match 10.7%; Score 73; DB 2; Length 463;  
 Best Local Similarity 24.3%; Pred. No. 17;  
 Matches 35; Conservative 18; Mismatches 55; Indels 36; Gaps 8;

OY 1 MASASVRLGVPGRMTDPRGI-----TEDEGRTWYV-----VRF 39  
 DB 12 LAGIAYVGLGAGAKLMLARQVAQETDELIAAFDLIEVPGFSSPALANGGCPVEL 71  
 OY 40 NPSREWARASGSKRYEPSTIVHLMQAVHTELLSSGMPFSQLPAYWQLYPKR---K 95  
 DB 72 RAKGEWLRVREFNRLDEPTIHW---GIRLPTMDGVYISOPPV---QPGSEFLYQ 124  
 OY 96 YRADP-SFWEIADHGQIDSMEL 118  
 DB 125 FKTDAGSTWY---HPHLMSEQL 145

RESULT 12  
 O9FI99 PRELIMINARY; PRT; 464 AA.  
 AC O9FI99;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GLUCORONOSYL TRANSFERASE-LIKE PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLDDB1;  
 RX MEDLINE=9937451; PubMed=10470850;  
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,  
 Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.  
 Sequence features of the regions of 1,011,550 bp covered by seventeen  
 P1 and TAC clones."  
 RL DNA Res. 6:183-195(1999).  
 DR EMBL: AB017060; BAB10792.1;  
 DR InterPro: IPR002213; UDPGT.  
 DR Pfam: PF00201; UDPGT; 1.  
 KW transferase.  
 SQ SEQUENCE 464 AA; 52360 MW; 522E2121P643896D CRC64;

Query Match 10.7%; Score 73; DB 10; Length 464;  
 Best Local Similarity 23.8%; Pred. No. 17;  
 Matches 31; Conservative 18; Mismatches 49; Indels 32; Gaps 8;

OY 16 LMIORPGIYDEDEGRTWY-----VYREMPSRREKARASGSKRYEPSTIV 61  
 DB 303 LMVVRPQ---SVHGDMTIESLPQFMSLIDGKIVAPQLDVLHRAFTG---FLTH 355  
 OY 62 HLMQAVHTELLSSGMPFSQLPAYWQLYPKRYRAADSFWELDH--GGID--SMEQ 117  
 DB 336 NGMN---STLESICG-VPMICLPCKMDQFVNAR---ISVWRVGIHLEGRIETRETER 408  
 OY 118 LVITYGPERK 127  
 DB 409 AVIRLWESK 418

RESULT 13  
 O9L803 PRELIMINARY; PRT; 513 AA.  
 AC O9L803;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PUTATIVE ACYL-CoA DEHYDROGENASE.  
 GN PTOFRFO.

OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 RN NCB1\_TaxID=316;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KC;  
 RA Lewis T.A., Cortese M.S., Sebat J.L., Green T.L., Crawford R.L.;  
 RT "Identification of a Region of the Pseudomonas stutzeri strain KC  
 Chromosome Containing Genes for the Biosynthesis of Pyridine-2,6-  
 bis(thiocarboxylic acid), the Agent of Carbon Tetrachloride  
 RT Dechlorination Produced by this Organism."  
 RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF196567; AAF3139.1;  
 DR InterPro: IPR001552; Acyl-CoA\_dh.  
 DR Pfam: PF00441; Acyl-CoA\_dh; 1.  
 DR PROSITE: PS00072; ACYL\_COA\_DH\_1; UNKNOWN\_1.  
 DR PROSITE: PS00073; ACYL\_COA\_DH\_2; 1.  
 SQ SEQUENCE 513 AA; 57432 MW; 4FD6E16F30D5C34 CRC64;

Query Match 10.7%; Score 73; DB 2; Length 513;  
 Best Local Similarity 24.3%; Pred. No. 19;  
 Matches 25; Conservative 16; Mismatches 42; Indels 20; Gaps 4;  
 QY 39 FNPSPREMARASQSGSR-----YESP-----ITVHMQMAYVTRFLLSSGMP 80  
 DB 45 FGPRTSWHSRSGAGRLVHCHEFORATCARRRRQVVSCHLHLLHPRRRARGVP 104  
 QY 81 FSLPAYMOLYR-GRKYRADS-SFWEIADHGQIDSMQVLT 121  
 DB 105 LGQLPRKQIDPSGRSLSLACTPEANDMDFSLRVQMLYST 147

RESULT 14  
 Q9EMP5 PRELIMINARY; PRT; 2229 AA.  
 AC Q9EMP5;  
 DT 01-MAR-2001 (TREMBLrel, 16, Created)  
 DT 01-MAR-2001 (TREMBLrel, 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)  
 DE PUTATIVE NON-RIBOSOMAL PEPTIDE SYNTHASE.  
 GN SC4C2.17.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN NCB1\_TaxID=1902;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb streptomyces coelicolor A3(2) chromosome."  
 RT Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL450432; CAC17499.1;  
 DR InterPro: IPR000873; AMP\_bind.  
 DR InterPro: IPR001242; DUF4.  
 DR InterPro: IPR003880; Phosphopant\_attch.  
 DR Pfam: PF00501; AMP-binding; 2.  
 DR Pfam: PF00668; Condensation; 3.  
 DR Pfam: PF00550; pp-binding; 3.

DR PROSITE: PS00075; ACP\_DOMAIN; 3.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
 KW Phosphopantetheine.  
 SQ SEQUENCE 2229 AA; 239295 MW; 66EB782247C326CD CRC64;

Query Match 10.7%; Score 73; DB 2; Length 2229;  
 Best Local Similarity 28.4%; Pred. No. 1e+02;  
 Matches 25; Conservative 6; Mismatches 29; Indels 28; Gaps 4;  
 QY 6 SVRLQRP-----PRLMIORPGL---YDEEGRTWTVVVR-NP 41  
 DB 878 SVYGVPLNMRARVYDGRGRCRDPDLPCELMIGSPGANGYNGDRGT---AERVVDH 933  
 QY 42 SRREMARASQSGSRPEPTVHLMQMAVH 69  
 DB 934 DGERWYRSGDLARYRDPGVLEFLGRADH 961

RESULT 15  
 O69839  
 ID O69839 PRELIMINARY; PRT; 513 AA.  
 AC O69839;  
 DT 01-AUG-1998 (TREMBLrel, 07, Created)  
 DT 01-AUG-1998 (TREMBLrel, 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)  
 DE PUTATIVE IRON-SULFUR BINDING OXIDOREDUCTASE.  
 GN SCIB5.11C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN NCB1\_TaxID=1902;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Harris D., Taylor K.;  
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb streptomyces coelicolor A3(2) chromosome."  
 RT Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL023517; CAA18965.1; -;  
 DR HSSP: P08980; IRFS.  
 DR InterPro: IPR00205; NAD\_binding.  
 DR InterPro: IPR001281; Rieske.  
 DR Pfam: PF00355; Rieske; 1.  
 DR PRINTS: PR00162; RIESKE.  
 DR PROSITE: PS00199; RIESKE\_1; 1.  
 DR PROSITE: PS00200; RIESKE\_2; UNKNOWN\_1.  
 SQ SEQUENCE 513 AA; 54931 MW; 06F0967279C9F4AF CRC64;

Query Match 10.6%; Score 72.5; DB 2; Length 513;  
 Best Local Similarity 32.8%; Pred. No. 21;  
 Matches 19; Conservative 7; Mismatches 19; Indels 13; Gaps 2;

QY 11 VPRGLMIORPG-----LYDEGR-----TWYTVVYRFRPSPREMARASQSGSR 55  
 DB 436 LPREGAVVRAAGRLAVYRDEGALHAVSPCTHGLCLVDFAAERARWCPCHGSRF 493

Search completed: November 29, 2001, 04:07:21

Fri Nov 30 11:03:13 2001

Job time: 222 sec

us-09-526-329-39.rspt

